Government of Ontario IT Standard (GO-ITS)

GO-ITS Number 56.4

OPS Business Intelligence and Business Analytics Reference Model

Version #: 2.0

Status: Approved

Prepared under the delegated authority of the Management Board of Cabinet

 Last Review Date: 2014-08-21

Copyright & Disclaimer

For third party users, including government contractors and entities seeking to provide products or services to the Government of Ontario, the Government of Ontario does not represent or warrant to you the accuracy, suitability or completeness of the content of this document.

Document History

Date	Summary
2014-08-21	**Approved:** • IT Executive Leadership Council approval • Clarification of terms: Pseudonymization, depersonalization, tokenization, anonymization added to Glossary • Approved version number set to version 2.0
2014-08-20	**Endorsed:** • Architecture Review Board endorsement • Standard renamed 'OPS Business Intelligence and Business Analytics Reference Model'
2009-09-24	**Approved:** • Architecture Review Board approval • Approved version number set to version 1.0
2009-06-25	**Created**: GO-ITS 56.4 OPS Business Intelligence Reference Model.

Requirements Levels

Within this document, certain wording conventions are followed. There are precise requirements and obligations associated with the following terms:

Must	This word, or the terms "REQUIRED" or "SHALL", means that the statement is an absolute requirement.
Should	This word, or the adjective "RECOMMENDED", means that there may exist valid reasons in particular circumstances to ignore the recommendation, but the full implications (e.g., business functionality, security, cost) must be understood and carefully weighed before choosing a different course.

Table of Contents

1. Foreword

Government of Ontario Information Technology (GO IT) Standards are the official publications on the IT standards adopted through the Office of the Corporate Chief Information Officer (OCCIO) for use across the government's IT infrastructure.

These publications support the responsibilities of the Ministry of Government Services (MGS) for coordinating standardization of Information & Information Technology (I&IT) in the Government of Ontario.

In particular, GO IT Standards describe where the application of an IT standard is mandatory and specify any qualifications governing the implementation of the IT standards.

1.1. Applicability

All ministries and I&IT clusters are subject to GO IT Standards.

All advisory and adjudicative agencies are subject to GO IT Standards. For the purposes of this document, any reference to ministries or the Government includes applicable agencies.

All other agencies that are using Ontario Public Service (OPS) information and information technology products or services are required to comply with GO IT Standards if they are subject to either the *Management and Use of I & IT Directive* or GO IT Standards by Memorandum of Understanding.

As new GO IT standards are approved, they are deemed mandatory on a go-forward basis meaning at the next available project development or procurement opportunity as applicable.

When implementing or adopting any GO IT Standards, ministries, I&IT Clusters and applicable agencies must follow their organization's pre-approved policies and practices for relevance of this standard and ensuring that adequate change control, change management and risk mitigation mechanisms are in place and employed.

2. Introduction

2.1. Background and Rationale

2.1.1. Preface

This reference model provides an abstract representation of the scope, key and important concepts, definitions, entities, functions, processes, components and their interrelationships that are related or relevant to the Business Intelligence (BI) and Business Analytics (Analytics) in the context of government programs and services to support adoption or expansion of a BI or Analytics program.

In the OPS many organizations use (or require) intelligence on their business to better meet the needs of Ontario citizens. Moving to a business intelligence and business analytics business model within an overall data management framework is a complex business transformation that can be supported by automated technology tools. Acquiring and implementing such solutions can be an expensive proposition requiring transformation initiatives to go through complicated, costly and time-consuming procurement, solution design/development, and governance (including financial and architecture oversight).

An environmental scan of BI and Analytics initiatives across the OPS found a wide variation in understanding by project teams and the resultant business architectural modeling and design.

The desired outcome of defining a BI and Analytics business model via this document is to ensure that required investments in technology tools or solutions are cost-effective, appropriate to meet business need, available when needed, and standardized and reusable where possible.

2.1.2. Objectives

The purpose of this document is to provide standardized definitions for use across the OPS, a common understanding of Business Intelligence and Business Analytics, and common entities that should contribute to the development of a BI and Analytics program. The document is intended to help equip readers to:

- Gain a common understanding of the generic business functions, roles, concepts, common framework, common terminology and standardized definitions needed to support a sound BI or Analytics program;
- Recognize whether they are in this BI and Analytics space;
- Understand how BI and Analytics could benefit the business and support the business case for developing a BI or Analytics program.

2.1.3. Mission Statement

Business Intelligence (BI) and Business Analytics (Analytics) are both widespread business needs that are often understood differently across the OPS. BI and Analytics programs contribute to business success by enabling businesses to develop new datasets and performance indicators, integrate disparate datasets, and perform dynamic analysis.

A well-managed BI and Analytics program provides (1) better and more responsive decision support as a result of more effective access to data, and (2) lower operating costs as a result of more efficient use of data infrastructure and reduced redundancy of data, hardware and technology administration.

To ensure strategic alignment between business and IT, every Business Intelligence or Business Analytics program should begin with a validated business-based mission statement such as, for example, the following:

"*The mission of the BI and Analytics program is the empowerment of business improvement opportunities for its sponsors through the delivery of information to the decision support community reflective of business processes and outcomes with optimal level(s) of formatting, timeliness, history, detail and quality to provide a reliable foundation for targeted business improvements*."[1]

2.1.4. Key Drivers

There are many reasons why Business Intelligence and Business Analytics would be considered important to organizations. For example, organizations may use BI and Analytics initiatives to reduce costs by implementing better fraud detection or developing more targeted policies and programs through better recognition of consumer segmentation. With the proliferation of mobile devices, the consumerization trend and self-service BI approach have enabled more users access to user friendly and easy to consume BI and Analytics tools, alerts, key metrics and performance indicators.

Growing Availability of Data: Digital data is everywhere – captured and created through an increasingly interconnected myriad of mobile devices, social networks, and client tracking systems. Integrating large datasets can reveal new insights or result in misinterpretation of patterns. So-called Big Data reflects this growing collection of

[1] Adapted from "Aligning BI with Business Strategy: How a Mission Mapped Architecture can help", Bair, J., S. Fox, M. Hunt, and D. Meers, DSSResources.COM, 09/10/2005.

datasets so large and complex that it becomes difficult to process using traditional data processing applications.

While global movements like Open Data reflect the growing recognition that the wealth of data and information available has the potential to drive innovation, foster greater transparency, and spur economic growth, there is also uncertainty around loss of control, security and privacy issues, and the risk that others may interpret government data with unintended results.

Need for Evidence-Based Decision-Making: Public sector organizations all over the world are under increasing pressure to evaluate and provide concrete evidence on whether their policies and programs are effective and efficient at meeting intended objectives. Evidence-based decision-making refers to the systematic and rational approach to researching and analysing available evidence to inform the policy making process.[2] Data driven evidence-based decision-making was a persistent theme in the 2012 report from the Commission on the Reform of Ontario's Public Service. While commission chair and economist Don Drummond highlighted the need for evidence to drive policy, he also cautioned that empirical evidence may present a more complicated picture that requires a deeper understanding of the available data.

2.2. Context

The relationship between BI and Analytics is not always clear. Business may view BI and Analytics as more of a continuum from the relatively straightforward (e.g. operational performance reporting) to the highly complex (e.g. advanced analytics to inform strategic decision-making) as depicted in Figure 1.

Figure 1: Continuum of Business Intelligence and Business Analytics

[2] Australian Bureau of Statistics Retrieved 2014-01-26 from http://www.abs.gov.au/ausstats/abs@.nsf/lookup/1500.0chapter32010

According to The Data Warehousing Institute (TDWI), BI and Analytics are similar in their use of data as measures, use of metrics to evaluate business behaviours, and the shared goals of understanding, insight, and foresight. Where they differ is in the way data is used and information created.[3]

Typical to Business Intelligence	Common to BI and Analytics	Typical to Business Analytics
Strongly oriented toward querying, reporting, OLAP, and alerts.	Use of data as measures.	Works with data to support business planning. Planning is more dynamic.
Answers questions such as what happened, how much, how often, and where the problems occur.	Use of metrics to evaluate business behaviours.	Answers complex questions about why things happened, the probability of events or conditions occurring under certain conditions, and what is realistic to expect for the future.
Typically produce scorecards and dashboards, suited to business monitoring with relatively static data structures and regularly published information.	Shared goals of understanding, insight, and foresight.	Users work constantly with unanticipated questions and other uncertainties.

BI and Analytics enable organizations to see where their business has been, understand where it is now and predict where it is going.

BI and Analytics facilitate the identification of potential cost savings through more efficient service delivery, explain why an event happened so organizations can take appropriate risk management actions, and identify new trends and opportunities with the insight from high-quality, well-managed data.

2.2.1. Data Management Perspectives

There are different Data Management perspectives from which to consider the challenges of accessing and managing the data in an organization. Efficient data access can only take place with the proper administration and support across the different Data Management perspectives as shown in Figure 2.

[3] TDWI Business Analytics: Exploration, Experimentation, and Discovery, 2012

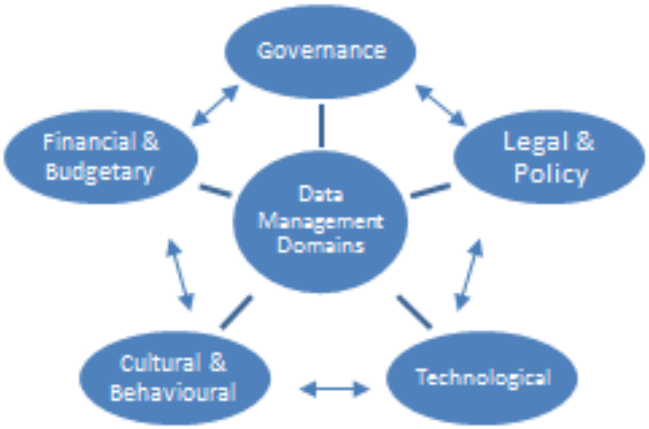

Figure 2: Data Management Perspectives[4]

The different data management perspectives consist of:

- **Governance** – to ensure transparency, responsibility and accountability
- **Financial and budgetary** – to ensure operational efficiency / effective data management to maximize reuse of data and avoid unnecessary duplication of data collection
- **Legal and policy** – to ensure the balance of intellectual property rights versus public good, lawful data management, respecting security, privacy and trade secrets
- **Cultural and behavioural** – to ensure quality of trusting that data are what they purport to be, professionalism building on codes of conduct and flexibility to allow different models of access
- **Technological** – to ensure interoperability of protocols, access and multiple use, quality including technical components of authenticity, integrity and security of data

2.2.2. Technological Perspective of Data Management

The complete scope of the Data Management discipline is significantly broader than Business Intelligence and Analytics. The technological perspective of Data Management, as depicted in Figure 3 encompasses the activities that are required to measure, track, monitor, maintain, support, use and protect an organization's data assets.

[4] Adapted from Data Science Journal, Arzberger et al, 2004

Figure 3: Technological Perspective on Data Management Discipline[5]

Business Intelligence and Business Analytics represent just a portion of the overall Data Management discipline with linkages and dependencies across different functions, people, processes and technologies. The red outlines in Figure 3 indicate the primary areas of focus related to this BI and Analytics reference model. While strong linkages and dependencies exist to other areas, the depth required to cover the remaining areas is unfortunately beyond the scope of this reference model.

2.3. Target Audience

The reference model is intended for the following target audiences:

Audience	Use of the Document
Business Planners, Policy Analysts, Business Analysts	• Supports business transformation or program review of government programs and services involving BI and Analytics.
Program Managers	• Encourages a focus on client needs and delivering value. • Supports business model identification and design for BI and Analytics.
Architects & System Developers	• Ensures that BI and Analytics business models are well formed to support alignment of information systems that have BI and Analytics requirements.

[5] Data Management Association (DAMA) Data Management Body of Knowledge (DMBOK), 2009

2.4. In Scope / Out of Scope

A BI and Analytics program must consider everything from after the data is recorded to when the information is presented for action or decision-making.

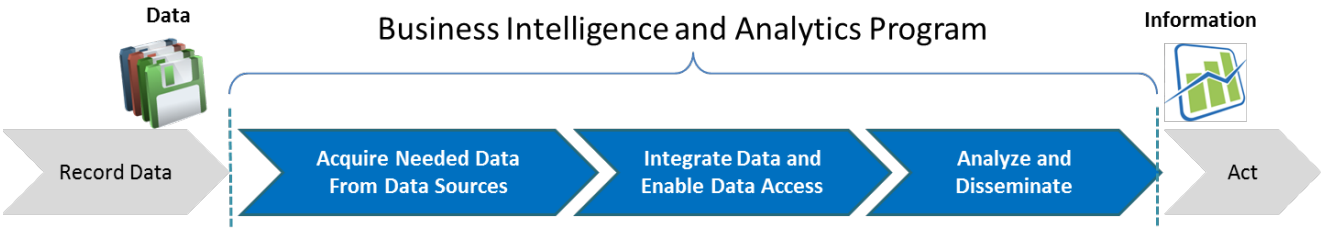

Figure 4: Scope of a BI and Analytics Program

For business users, BI and Analytics represent an interface to data, with much of the technical complexity hidden but no less important to the end-to-end BI and Analytics program. The scope of the overall BI and Analytics technology stack includes:

- Data rationalization technologies such as data integration (e.g. extract, transform and load), data quality, data profiling and master data management (MDM);
- Derived data technologies, where data has been processed or transformed from its original raw form, e.g. staging, operational data store, data warehousing, data marts, and cubes;
- Data usage and delivery technologies such as reporting and querying, text and content analytics, visualization, and agile search and discovery.

In the context of this reference model, BI and Analytics are often referred to together to reflect their common aspects. Where possible, common business functions, processes and services are described. However, where there are differences between BI and Analytics the differences will be described.

In Scope	Out-of-Scope
Identification of common business functions of BI and Analytics that can be represented generically	Identification of business functions that support the delivery of a program using data but are not specific to business intelligence or analytics (e.g. transaction based approaches).
Selected OPS business architecture artefacts to describe a generic BI/Analytics business model. Note: Selected artefacts must be assessed for relevance and completed for a particular ministry program implementation.	An end-to-end business architecture for BI and Analytics including all artefacts. Information/data model to support BI/Analytics. Specific information requirements will be business specific.

Types of business rules to consider for a BI and Analytics business model.	Business rule source instances or business rule statement instances (i.e., business specific).
Identification of linkages to Data Management functions and domains of significance to BI and Analytics within the OPS, such as master data management and reference data. Statement of Data Management principles related to standardization / definitions.	The complete scope of the Data Management discipline or comprehensive details on functions not specific to business intelligence or analytics (e.g. document and content management). Defining or implementing any business specific data standards or mandatory requirement for conformed dimensions.
	Prescriptive reference architecture for BI and Analytics to be used with specific products or solutions.
	Cloud computing challenges such as commingled data, cloud data ownership, cloud security policy, etc.

3. Technical Specification

3.1. Guiding Principles

The following are core information management principles that should govern the development of any BI and Analytics program:

- **Data as an asset** – Data should be treated as a valuable asset and managed appropriately to ensure it is available to help an organization achieve its strategic objectives. As an asset, Data can be identified, described, transformed, collected, and invested in. It can also grow stale, incur costs or increase in value when linked with other data to create new information and insights.
- **Access to data** – Business should be able to access and integrate high quality data from multiple sources in a single secure environment to meet current and emerging program needs, enable research of innovative effectiveness questions, and eliminate the need for labour-intensive manual intervention to combine and cleanse data from disparate sources.
- **Collect once, use many times** – Solutions should reduce redundant data collection processes, manipulation and storage to reduce costs, avoid wasted time, labor and productivity, avoid many data quality issues, and improve system performance.
- **Openness and transparency** – Publically funded data should be openly available to the maximum extent possible except where restricted for legitimate reasons e.g., health & safety, protection of privacy, etc.
- **Privacy by design** – Privacy analysis must be conducted whenever personal information is involved. When required, a mechanism must be available to de-identify personal information.
- **Scalability** – Solutions should be designed to support the anticipated volume of data and activity.
- **Self-service** – Business should be provided with timely, appropriate access to the data, and be trained on appropriate tools, so that they may effectively conduct innovative analyses.
- **Supportability** – Solutions should use standardized, modern, vendor-supported hardware and software that can easily be supported by the organization.

3.1.1. Open Data Requirements

Use of BI tools and services may involve mashing-up and combining data from multiple sources. This may result in the creation of a new data asset. Such new data asset should be accessible, efficiently used and open when deemed appropriate by the program area.

To ensure that the principles of access to data, collect once use many times, and openness and transparency are part of the technical implementation of a BI and Analytics solution, program areas and IT clusters **must** ensure that all BI and Analytics applications procured or developed include a function to export the source data used for a specific analysis into an open standard file format such as CSV, XML, or TXT.

Exporting the associated data into a vendor neutral file format will allow business areas to guarantee access to specific data sets beyond the limits of a specific contract or application, ensure that the data supporting a specific BI and Analytics process can be re-used by other program areas using other applications, and allow for quick integration between BI and Analytics tools and the Ontario Open Data Catalogue.

3.1.2. Definitions and Characteristics

The following are OPS adopted definitions of the key concepts relevant to this reference model. Along with each definition are contextual descriptions to expand on how the concepts should be understood from their respective business and technology perspectives.

Term	Definition and Contextual Descriptions
Business Intelligence (BI)	DEFINITION: Business Intelligence is the transformation of raw data into meaningful and useful information used to enable more effective strategic, tactical and operational insights and decision-making that contributes to improving overall enterprise performance.[6] Enterprise BI Reporting solutions inform decision-making based on historical data by collecting data from a variety of sources and creating specific reports using pre-selected criteria and consistent metrics. Technologies that support Business Intelligence include tools for reporting, dashboards, ad hoc queries, cubes, online analytical processing, scorecards, graphs, desktop tools integration, GIS integration, and mobile BI.
Business Analytics (Analytics)	DEFINITION: Business Analytics is the identification of meaningful patterns and correlations among variables in complex, structured and unstructured, historical, and potential future datasets for the purposes of predicting future events and assessing the attractiveness of various courses of action with the goal of highlighting useful information, suggesting conclusions and supporting decision-making.[7]

[6] Adapted from Forrester, “Craft Your Future State BI Reference Architecture”. November 2012. Evelson and Yuhanna.
[7] Updated definition based on Gartner, Forrester and GO-ITS 56.4 BI Reference Model version 1.0, 2009

Term	Definition and Contextual Descriptions
	The discipline is focused on understanding business performance based on deep statistical and quantitative analysis, data mining, econometrics, forecasting, text analytics, explanatory, prescriptive and predictive modeling and fact-based and integrated risk management. Technologies that support Business Analytics include tools for statistical modeling (predictive, regression, clustering, and association), agile search and discovery, interactive visualization, sentiment analysis, structured and unstructured data analysis.

Table 1: OPS Definition of Key Concepts

Business Intelligence and business analytics should provide the most comprehensive decision support and data analysis capabilities available via a broad range of tools and technologies such as interactive dashboards, ad hoc analysis, proactive detection and alerts, advanced reporting, statistical modeling, and visualization.
BI and Analytics business models include the following common characteristics:

- **Pervasive** – Deliver intuitive, role-based intelligence or analytics capability from front line employees to senior management that enable better decisions, actions, and business processes.
- **Comprehensive** – Integrate data from various domains such as financial performance management, operational intelligence, and transactional applications.
- **Reuse** – Leverage existing data sources and systems.

A business initiative that has these characteristics qualifies as a business intelligence and/or business analytics initiative. Business Intelligence and business analytics are not defined by the functionality of a supporting tool that must be configured to support the "to be" business intelligence business model.

3.2. Business Architecture

This section contains selected Business Architecture artefacts that describe a notional Business Intelligence and/or Business Analytics business model which is presented at a high level to be used as the foundation for any BI and/or Analytics initiative in the OPS. Descriptions of these artefacts can be found in the GO-ITS 56 OPS EA Review Requirements Guidebook (https://www.ontario.ca/government/information-technology-standards).

3.2.1. Semantic Model

The following semantic model establishes the vocabulary that will be used by the enterprise to describe the business of a notional BI and Analytics program. The model is read by relating each entity to the connecting action in the direction of the arrow.

For example, the top of the model should be read as Business Intelligence, Data Delivery and Business Analytics *supports* Evidence-Based Decision-making and Open Data.

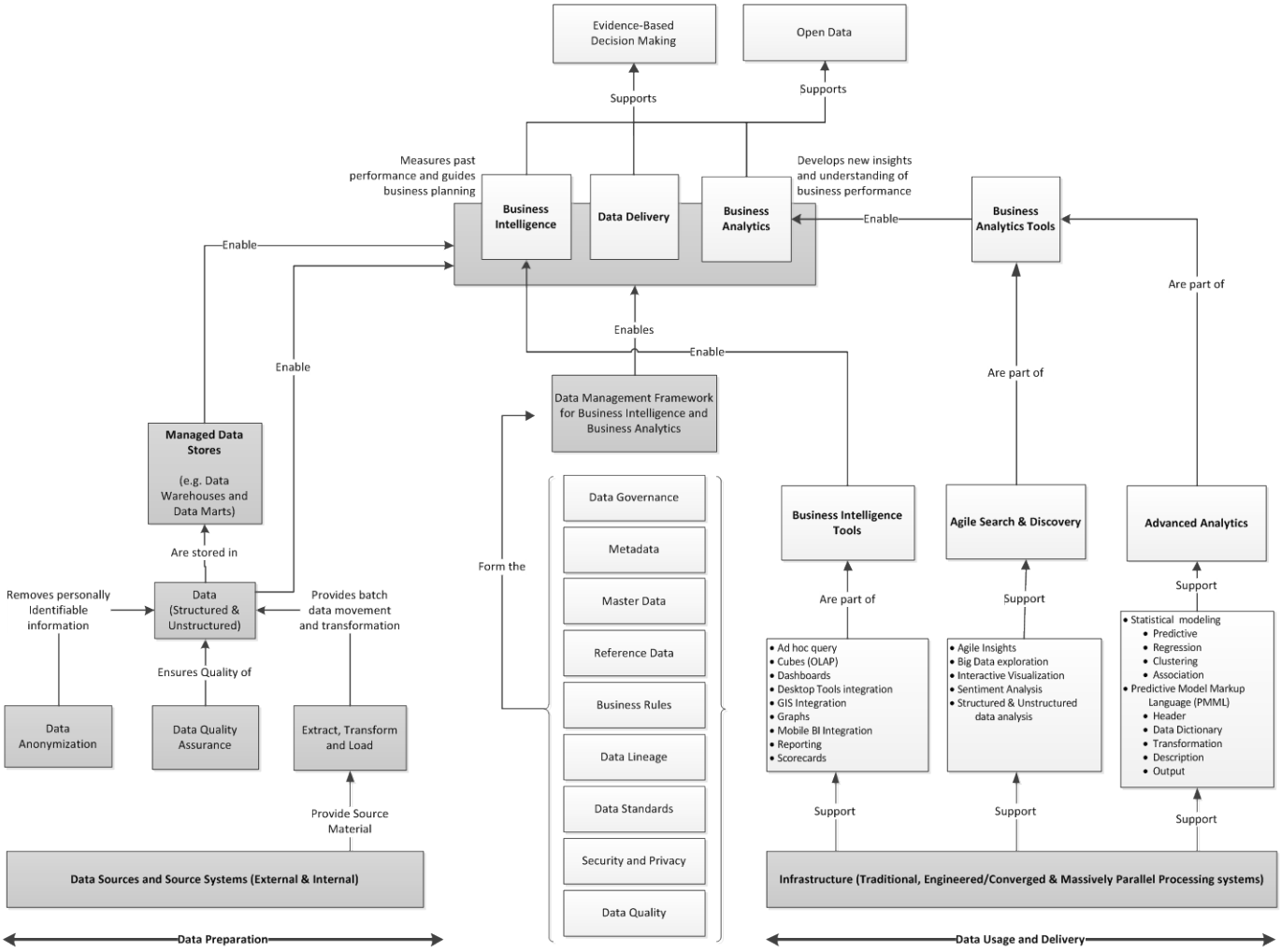

Figure 5: Semantic Model for the notional BI and Analytics Program

The lower portion of the semantic model in Figure 5 is divided roughly into three major sections, all enabling Business Intelligence and Business Analytics. The following describes the sections from left to right:

1) ***Data Preparation*** begins with the data sources and source systems (external and internal) that provide source material to an Extract, Transform and Load (ETL) process. The ETL process provides batch data movement and transformation for the data to be stored in Managed Data Stores (Data Repositories) such as Data Warehouses, Data Marts, and Operational Data Stores. It is important to note that while Data Collection systems are not strictly within the scope of BI and Analytics, they play an important role in providing data sources and maintaining source systems. Data Quality Assurance ensures the quality of the data is fit-for-purpose and Data Anonymization removes personally identifiable information from the data.

2) ***Data Management Framework for BI and Analytics*** is an important pillar to enable BI and Analytics, including Data Delivery. It is formed from the management of Data Governance, Metadata, Master Data, Reference Data, Business Rules, Data Lineage, Data Standards, Security and Privacy, and Data Quality.

3) ***Data Usage and Delivery*** is comprised of infrastructure systems that support the resources and capabilities that ultimately enable Business Intelligence and Business Analytics.

   Resources such as tools for reporting, dashboards, ad hoc queries, cubes, scorecards, graphs, desktop tools integration, GIS integration and Mobile BI integration *are part of* business intelligence tools which *enable* Business Intelligence. Agile insights, Big Data exploration, interactive visualization, sentiment analysis and structured and unstructured data analysis *support* Agile Search and Discovery, while Statistical modeling and Predictive Model Markup Language *support* Advanced Analytics, and together they *are part of* business analytics tools which *enable* Business Analytics.

3.2.2. Business Function Model

The following Business Function Model (BFM) in Figure 6 shows the business functions that support a Business Intelligence and Business Analytics Program. Business Function Models are defined in GO-ITS 56.[8] Following the BFM, a description of each business function is provided. This BFM is a re-useable model that should be leveraged by OPS practitioners designing or amending a BI and Analytics program. Justification must be provided if these functions are not included in the program design. For simplicity, each business function includes managing standards and processes, as well as their execution (e.g. it includes planning, provisioning and delivery). In more detailed modeling for a program, it may be desirable to separate them.

[8] GO-ITS 56 Appendix B

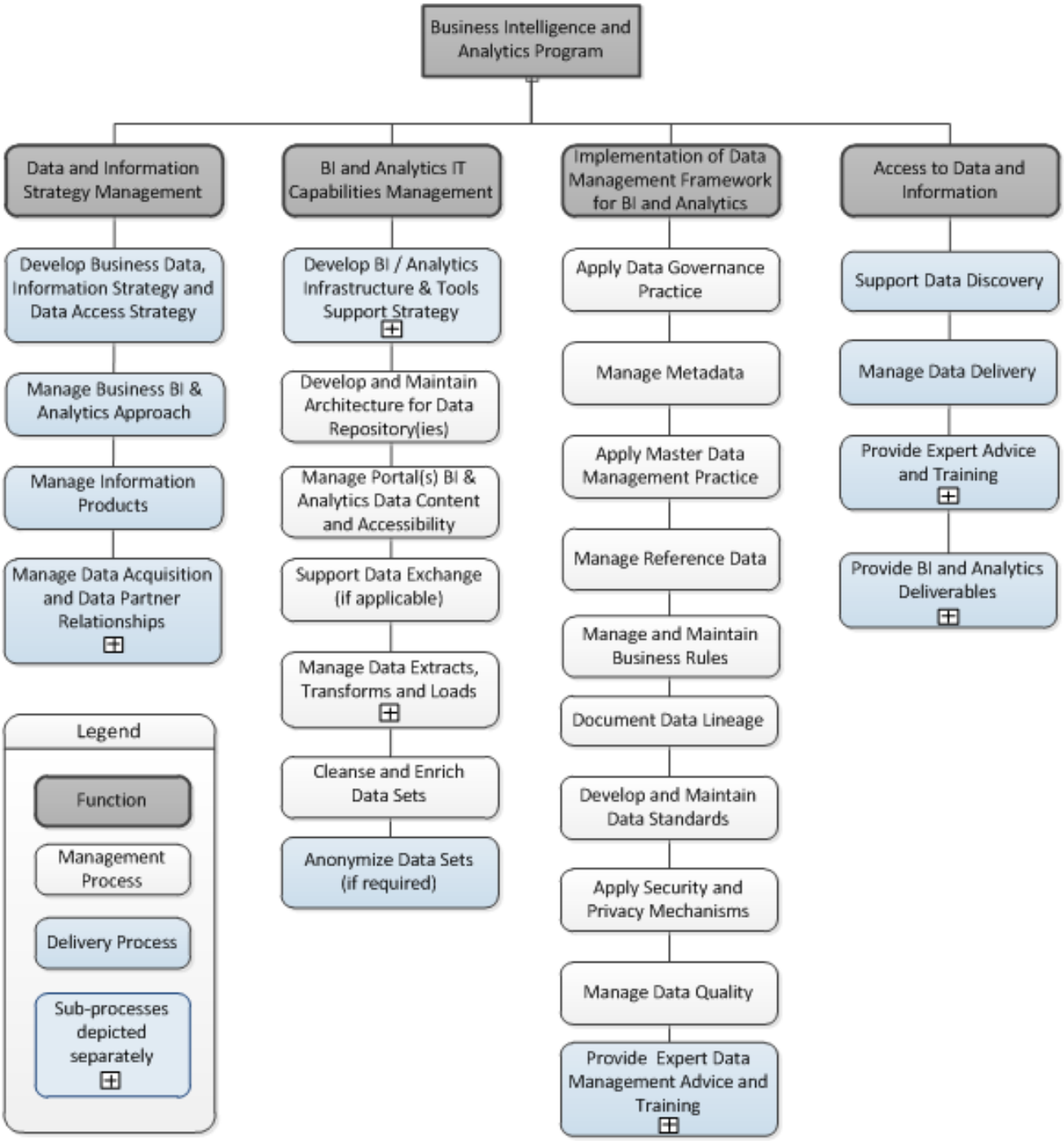

Figure 6: BI and Analytics Business Function Model

It is expected that possible sub-processes[9] not depicted in Figure 6 may be developed in further detail as appropriate. Readers should refer to other complementary guidance documents or communities of practice for examples of potential sub-processes.

[9] NOTE (from GO-ITS 56 Appendix B): Management processes are internal and administer the resources and operations of the organization. Delivery processes are client facing and supply the outputs of services.

Business Architecture practitioners are encouraged to group and modify these activities/processes to suit their business needs. It is not the intention of this document to suggest each process will necessarily map to its own unique business process model. Instead, the processes represent activities and tasks that must be performed in order to complete the desired business function.
In addition to the business functions outlined under a Business Intelligence and Business Analytics Program, there are business functions that belong to other Programs, Services, or Government of Ontario IT Standards such as the following that must be considered and aligned where applicable:

Open Data:

- All data should be considered for release as Open Data, which is the provision of the data sets, aggregated data and/or derived information products to the public through the OPS Open Data and Open Government initiative.

Information Management:

- Information Management Life Cycle applies to all data (structured and unstructured), and includes (but is not limited to) application of the OPS Information Security and Privacy Classification Policy (ISPC), and management of data retention schedules.

Enterprise Architecture (EA):

- Management of Enterprise Data Architecture (at conceptual, logical and physical) levels will rise in strategic importance to support a BI and Analytics program.
- Gathering of business requirements will require additional attention to data needs as the Fact and Dimension matrix must be created for a project that develops or acquires data warehouse and/or data mart based solutions for decision support initiatives.

Software Development Life Cycle (SDLC):

- Business and I&IT engagement and relationship management will include scope and linkages related to the Manage Data and Information Strategy function.
- Execution of technology solutions is directed by the Manage Data and Information function and Manage IT Capabilities function (includes new BI, GIS and Analytics tools, platforms, strategy, target architecture and solutions coordination), after appropriate governance.
- A typical IT transactional solution will involve the identification of business processes which will be supported by the solution. While this is desirable for BI and Analytics, it is not always applicable or necessary beyond their fit within policy or other decision-making. If business processes are developed, it is critical to identify the information and data interactions – where data is produced or required, and what it is.
- As part of determining business requirements for a BI solution, the business KPIs (Key Performance Indicators) to measure success must be identified and documented.

- In determining the solution design for BI and Analytics solutions, additional skills are needed, such as a strong understanding of the potential outputs (e.g. ad-hoc reports, dashboards, GIS, etc.), as well as data visualization skills to ensure that the information is presented so users can easily understand it. For Analytics tools, it is assumed that the statistical methods and models will be provided as business requirements.

Business Function Name	Related Business Processes	Description
Data and Information Strategy Management		Carry out the work to engage ministry business to determine their needs, develop strategies and execute overarching business processes to address them. Develop relationships with data providers that mature into formal agreements to acquire data. Consists of the engagement, planning and management to address business information needs, including fulfilment of ad hoc and ongoing needs.
	Develop Business Data, Information Strategy and Data Access Strategy	Identify the strategic data, data integrations, tool capabilities, etc. to meet business requirements. Includes identification of potential data aggregation, data sources and tool gaps. The Data Access Strategy describes how the ministry wishes to arrange for data requests and sharing, and prioritizing decisions to acquire new data (purchase, share, etc.) based on current utilization, business strategies and business key questions.
	Manage Business BI & Analytics Approach	Determine and manage how the business will approach BI and Analytics. For example, will analytics be limited to the users, shared locally or published broadly? If so, how will this work and what approvals or governance is required at each step?
	Manage Information Products	Execute Information Strategy by managing the available Information Products (e.g. dashboards, analytics models and scenarios, etc.) for the purposes of facilitating the delivery of decision support to consumers, including the provision of a searchable information products catalogue. If necessary, access to the catalogue will be in accordance with authorization and access controls (e.g. for sensitive scenarios or analysis or those involving personal information).

Business Function Name	Related Business Processes	Description
	Manage Data Acquisition and Data Partner Relationships	Execute Data Access Strategy by managing the formal processes and templates related to the acquisition of data, up to but not including the actual movement of data. Includes determining data sources, defining data source business rules, establishing accountabilities, determining data refresh schedules and managing data sharing agreements (negotiating provision of data, developing and managing formal data sharing agreements with data owners or data stewards). Arrangements can be for internal source data, as well as external data, data purchases and the management of the library of data sharing agreements. Overall objective is to manage the ongoing relationships to ensure the data provided meets requirements and the terms of data sharing agreements and to ensure incoming data can be understood / used appropriately, per the BI and Analytics program standards and processes set in other Functions, such as the documentation of Metadata and associated business rules (to the data element level).
BI & Analytics IT Capabilities Management		Manage creation and maintenance of IT capabilities and platform(s) to support the BI and Analytics program. Includes infrastructure and tools.
	Develop BI / Analytics Infrastructure & Tools Support Strategy	Manage platform and tools, including those for data movement (ETL) and the Data Exchange capability (if applicable). Includes designing and implementing the BI and Analytics environment, and defining and testing the information delivery mechanisms.
	Develop and Maintain Architecture for Data Repository(ies)	Manage the overall architecture for the storage and movement of data within and between data the repository(ies) to potentially serve multiple users and support multiple IT solutions.
	Manage Portal(s) BI & Analytics Data Content and Accessibility	Provide life cycle coordination and request execution of portal(s) for access to data, BI and Analytics. Portals are typically web-enabled and bring relevant evidence-based decision-making supports from diverse sources together in a uniform way.
	Support Data Exchange (if applicable)	Support the necessary processes, standards and content if a “publish-subscribe” model for data is in place e.g. between transactional databases, data repositories or other sources.

Business Function Name	Related Business Processes	Description
	Manage Data Extracts, Transforms and Loads (ETL)	Manage data moves and transformations to and within data repositories, to support direct data use and use by tools. Includes obtaining (a feed of) the source data from data providers, classifying data sets, developing harmonization processes, establishing data enrichment processes, developing data cleansing processes, developing data aggregation processes, documenting data traceability, developing data purging processes, managing data quality control processes, executing change management processes and developing data retention schedules. If source data is not harmonized when received, ETL can change the data format (data type, size) of the source data so that it conforms to agreed standards.
	Cleanse and Enrich Data Sets	Cleanse data by processing changes to the source data according to a set of processes to make sure it is correct and accurate. Sometimes called data scrubbing or data transformation. Enrich data by changing or enhancing the source data with additional data so that it enables linking with other data sets within the repository.
	Anonymize Data Sets (if required)	Change source data by applying the agreed upon anonymization method to minimize risk of re-identification and/or data breach.
Implementation of Data Management Framework for BI and Analytics		Implement the Data Management practices necessary to support a BI and Analytics Program. Business Architecture for the BI and Analytics Program must illustrate how the relevant components of the Data Management Framework will be incorporated, or explain why they are not applicable to the specific Program.
	Apply Data Governance Practice	Identify individual and organizational roles and responsibilities for evidence-based decision making, supported by the management and use of the data and information. The need for governance is identified in other functions, and could include for example: - Data Stewards assignment and changes. - Strategies endorsement (e.g. Data acquisition and IT solution priorities).
	Manage Metadata	Develop approach to collect, update and share Metadata, both to fulfill direct business needs (e.g. data definitions and documentation of data quality or lineage) and support technical metadata needs.

Business Function Name	Related Business Processes	Description
	Apply Master Data Management Practice	Develop approach to identify potential Master Data, bring it forward for Data Governance decision, and implement and re-use it.
	Manage Reference Data	Develop approach to maintain reference data values and associated metadata, and communicate any reference data additions or changes to reference data consumers.
	Manage and maintain Business Rules	Develop and execute coordinated approach to document, communicate and re-use business rules related to data.
	Document Data Lineage	Update Metadata data dictionary to ensure it documents how the particular data element was changed and provide a capability for the data consumer to trace the data back to the source dataset.
	Develop and maintain Data Standards	Establish and update or identify the data standards used within the program. For example, if the program involves multiple collections of similar data, then a standardized approach or format is desirable. Data standards may need to extend well outside the BI and Analytics program to ensure incoming data can be used. Note that some other Functions may result in the creation of a Data Standard (e.g. Manage Metadata).
	Apply Security and Privacy Mechanisms	Manage a coordinated central function which enforces security and privacy requirements, including the tracking of requests and approvals for access to data and information products.
	Manage Data Quality	Develop the data quality framework to be followed, including control measures, quality assurance, etc. Implement the applicable quality activities as data is acquired, goes through ETL (Extract, Transform, and Load) or integrated.
	Provide Expert Data Management Advice and Training	Provide training, consultation and expert advice on relevant aspects of Data Management Framework for BI and Analytics. Includes identifying business process requirements, information requirements, and Open Data requirements, defining business intelligence outputs, business analytics outputs, and Open Data outputs, identifying business value and key performance indicators, and developing and delivering education.

Business Function Name	Related Business Processes	Description
Access to Data and Information		Manage the access to data and information within the BI and Analytics program.
	Support Data Discovery	Define the channel for data discovery and how the data will be delivered to the consumer (e.g., on screen, mobile device, report, etc.). This feeds into the documentation of an information products catalogue.
	Manage Data Delivery	Manage delivery of publish datasets and deliverables (including Metadata and/or Business Rules, as required) to the defined environment managed in accordance with the agreed upon authorization and access controls. May also include management of feedback and continuous improvement.
	Provide Expert Advice and Training	Provide BI and Analytics training and expert advice (e.g. visualization, GIS, scenario modeling, statistics, etc.). This could include a wide range of topics (e.g., tools, business scenario, etc.) and could be delivered using various mechanisms (e.g., direct, online, self-service, etc.). Includes developing and delivering education, as well as resolving end user tooling issues. Training would also include business specific training (e.g., specific business scenarios) so that the business is trained to understand how to interpret the information being conveyed and how it may initiate business decision making or actions to be taken by them.
	Provide BI and Analytics Deliverables	Provide BI deliverables and carry out Analytics using existing tools and data (including ad hoc data integration for individual use or limited sharing, typically using Data Scientist skills, such as statistical modeling/visualization). Includes implementing security and privacy mechanisms, implementing data retention and archiving practices, issuing alerts, publishing performance reports and dashboards, and publishing data sets.

3.2.3. Context Diagram

The Context Model shows key interactions with external entities of the following types:

- ***Accountability*:** Authority and responsibility to the enterprise governors
- ***Services***: Access and fulfilment of in-scope services to the client (target group)
- ***Collaboration*:** Partnerships with shared accountability to ensure client needs

- ***Procurement:*** Supplier relationships, often dealing with service production and support

The context diagram for the notional BI and Analytics program states the target groups of their services, and the outputs those clients receive. In addition, it represents stakeholder partnerships (collaboration and procurement) that must exist to deliver programs and services in the OPS.

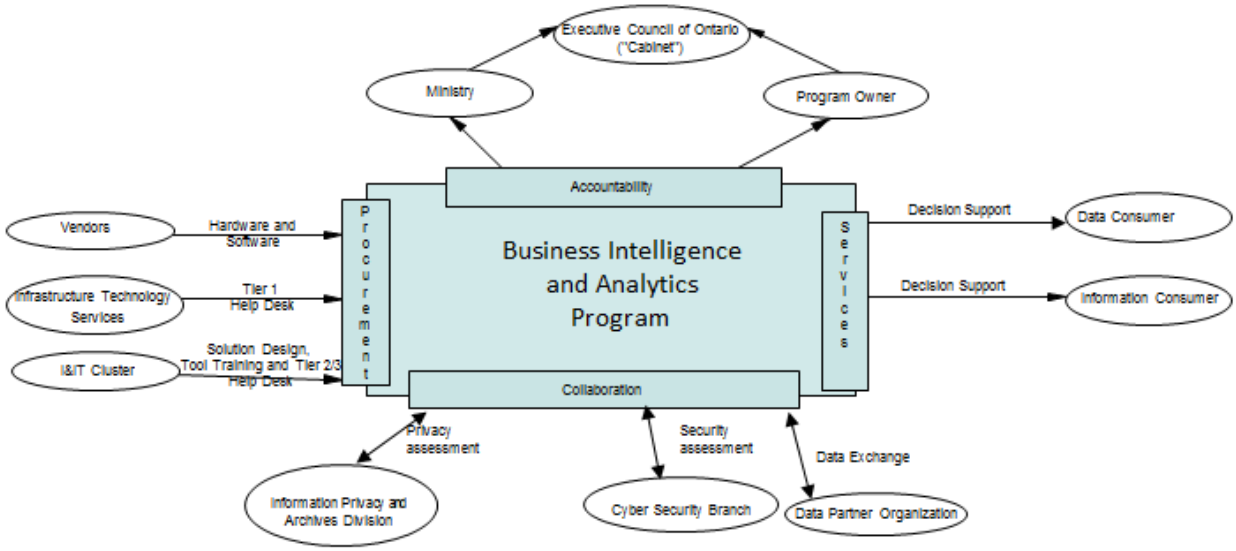

Figure 7: BI and Analytics Context Diagram

Modifications will be required to this context model to suit the unique service delivery owner/provider relationships that exist between and amongst OPS organizations.

3.2.4. Service Integration Accountability Model

The Service Integration Accountability Model (SIAM) is a diagram that shows how services are integrated in order to deliver the final deliverable to the end client (refer to GO-ITS 56 appendix B). The following SIAM represents the service offerings of a notional OPS BI & Analytics Program.

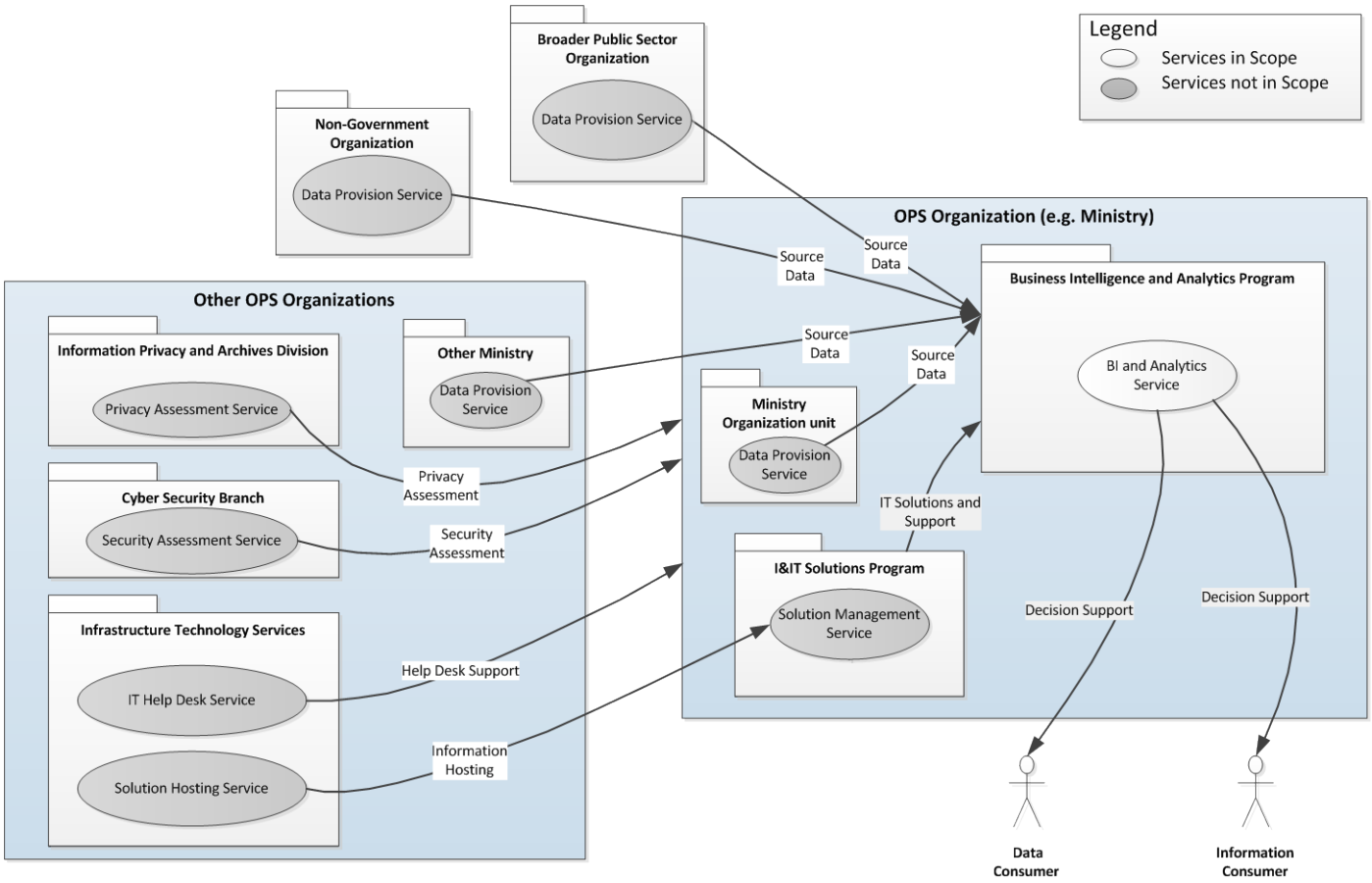

Figure 8: BI and Analytics Service Integration Accountability Model

The service deemed to be in scope for the notional BI and Analytics program is further described through a Service artefact. As a starting point for BI and Analytics program and projects, the SIAM focuses exclusively on the service that such a program is expected to deliver to its target groups (Data Consumers and Information Consumers).

3.2.5. BI and Analytics Service

Services provide specific results (service outputs) that satisfy the needs of a target group (i.e. client group) and contribute to the achievement of their program goals.

Business Architecture practitioners are expected to build upon the following service description to include the necessary details for service planning, provisioning and

delivery, including specific service delivery processes and mechanisms (e.g. delivery channels) available for their target group(s).

Service		
Type:	☐ **Public** ☒ **Support**	
Name:	BI and Analytics Service	
Description:	The provision of decision support in the way of data, information and/or information enablement (e.g. BI and Analytics tools) that meets the stated requirements of the Data and Information Consumer. Includes supportive advice regarding various mechanisms for translating data into timely, accurate, and useful information to support informed decision-making. The nature of the advice can be focused on datasets available for integration and/or the business intelligence and/or analytics resources that can be produced to suit the needs of the client group. Examples of Data Consumer decision support products may include Available Data Search; Data; Data Advice; and Data Discovery. Examples of Information Consumer decision support products may include Available Information Search; Information Advice (e.g. Modeling, Tools); and BI and Analytic Capacity	
Output:	Decision Support	
Output Type:	Advisory Encounter[10]	
Client Group:	**Name:**	**Description:**
	Data Consumer	An individual or organization within or outside the enterprise that obtains data from the BI and Analytics program.
	Information Consumer	An individual or organization within or outside the enterprise that obtains information from the BI and Analytics program.

3.2.6. Role Type

Role Type lists the types of roles played by parties (individuals and organization) of interest to the business. The following highlights key roles relevant to the notional BI and Analytics Program in the OPS. Additional examples of Roles can be found in the appendix.

[10] As per the GSRM, services that provide an encounter during which data, information and/or advice is conveyed to a party or system have a service output type of An Advisory Encounter. For this reason, all services offered by the notional BI/Analytics program share the same output type.

Role	Role Sub-Type	Description	Party or Parties playing the role
Data Custodian		An individual or department that is responsible for the safekeeping, transport, data storage and implementation of business rules, based on the applicable Data Sharing Agreements, instructions/decisions received from the Data Steward/governance body, and in line with direction provided by the GO-ITS standards or other OPS and Ontario rules/legislation.	E.g., Ontario Ministry Division, Ontario Ministry Branch
Data Owner		An individual or department that has the accountability for the data generated and/or consumed. Data owners are generally people who initially created or produced the data, and/or have authority and accountability under legislation, regulation, policy or other instrument for that particular data. The Data Owners make decisions on the data such as whether to provide data for disclosed uses, what safeguards should be in place to manage data risks, classification and access decisions regarding the collection, transformation, use, retention and disposal.	
Data Steward		An individual or department that serves as the appointed trustee for data assets and ensure the quality, accuracy and security of the data. Data Steward has the delegated responsibility to manage the Owner's data and its usage on the Owner's behalf. Where the Data Steward is not also the Data Owner, the Data Steward is bound by Data Sharing Agreement and/or other agreements and disclosures made with the Data Owner when the data was collected for use.	

3.3. Information Architecture

Information architecture describes all the things (terms, facts and concepts) and the relationships that are important to the enterprise (e.g. ministry, cluster) or a specific domain (e.g. an OPS program such as child protection program, a business function such as transfer payment management). It gives a formal representation of the data needed to run an enterprise or a business activity. It also represents the data structure and design of the in-scope information environment for an IT project or system.

Information architecture is primarily used to:

- Specify, analyze and represent business concepts and their relationships, and to facilitate common understanding and stakeholder agreement on the meaning of terms and relationships.
- Enhance communication with business staff and to clarify business rules involving the business information.
- Identify opportunities to share and/or integrate data.
- Enable information exchange and access between people and applications.
- Define data requirements, structure and design database for I&IT solutions.

From an Information Architecture perspective there are a number of considerations that need to be in place in the organization irrespective of the type of BI or Analytics solution being developed.

3.3.1. Data Modeling

Data modeling involves identifying things (entities) of importance for an organization, the properties (attributes) of those things and how the things are related to each other (relationships).

A data model can be used to describe the data or information requirements for different types of applications, and for general business understanding.

Data Modeling for Business Intelligence

While entity relationship modeling is a technique that seeks to remove the redundancy in data, dimensional modeling is a technique that facilitates data analysis (slicing and dicing and rolling up facts and measurements by dimensions). It presents the data in a format that allows for user understandability, query performance, and resilience to change. A dimensional model can be used to model the information requirements for a Business Intelligence solution such as a data warehouse or data mart.

Dimensional models are developed by leveraging the data definitions from the conceptual data models (CDM) and/or logical data models (LDM) of the operational data sources. **GO-ITS 56.3 Appendix A: Information Modeling Handbook (IMH)**

Chapter 3 provides the data modeling standard for the data models (e.g. fact and dimension matrix, warehouse models, dimensional models) required to support a decision support system within the OPS.

The OPS Business Architecture Concept Model (BACM) and the Common Data Elements Models (CDEM) for Party and Address provide a good starting point for defining conformed dimensions.

The BACM provides a high-level view of generic OPS business architecture concepts, terminologies, and their relationships, while the CDEM Party and Address models provide some required elemental data for data analysis. The following business analysis scenario illustrates the use of the BACM.

Business Analysis Scenario

Business Problem	- Ontarians are waiting too long to access emergency services
Usage of the BACM	- The business problem analysis process makes extensive use of data - The BACM helps frame the business problem, exploring the data model to find the right information and insights to respond to the problem. - Figure 9 illustrates how to use the BACM to discover and define the data elements for analyzing the business problem. For example: o To define the problem and identify the impact to the business: ▪ Use the PARTY ROLE subject to identify who will be impacted ▪ Use the PROGRAM SERVICE subject to identify which program(s) and services will be impacted ▪ Use the PEFORMANCE MEASURE subject to identify the impact to the business goals o To pose questions to the business: ▪ Use the data relationships which represent business rules to discover the relevant business questions o To develop the conceptual data model for the in-scope data requirements: ▪ Use the in-scope data elements and relationships identified in the BACM as a starting point, which can be further elaborated by adding next level details o To discover the facts and dimensions: ▪ The many-to-many relationships may be turned into Facts ▪ The descriptive entities may be turned into dimensions, such as LOCATION, PARTY ROLE, SERVICE, etc.

Business Problem:
•Ontarians are waiting too long to access emergency services

what does the business need to know?

Business Questions:
•Who are waiting for the emergency service, where and when?
•etc

what data are required to answer the questions?

Data required as measures:
•Average waiting time for the emergency service in the past five years by region and client age group, etc

what data to analyze?

Discover and Identify data elements and relationships:
•Waiting time, emergency service, region, client age group, etc

reusable data elements and relationships defined in the BACM

- **Geographic Location**:
  - Physical location (e.g. region)
- **Service**:
  - Service Type (e.g. public service)
  - Service Name (e.g. Emergency service)
- **Party**:
  - Party Role (e.g. client)
  - Party Type (e.g. individual)
    - Reusable individual attribute (e.g. Birth Date will be used to calculate the age and match to the client age group)

further usage examples

- Transform to fact-dimension matrix, and then the logical dimensional data model, etc.

Fact Group	Dimension				
	Time	Region	Service	Client	Age Group
Client Waiting time	X	X	X	X	X

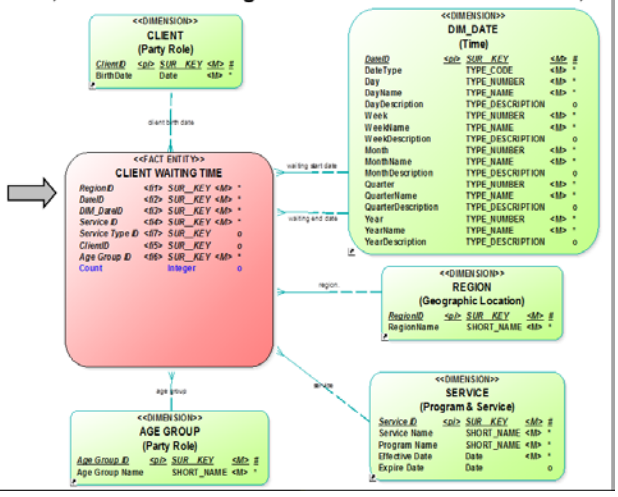

Figure 9: How to utilize the BACM to discover and define data elements

3.3.2. Data Standards

Data standards are an agreed upon set of rules on the representation, format, definition, structure, use and management of data. Data standards play a key role in enabling information sharing, data integration and reuse.

Programs must exercise due diligence in conducting searches for existing data standards prior to solution design or modernization. Begin by referring to https://www.ontario.ca/government/information-technology-standards

From an OPS perspective there are common standards that must be used consistently across the OPS and there are industry specific standards that are applicable to only a specific business area (e.g. Ministry of Health). Some examples:

- **Common Data Standards:** These are examples of some of the data standards that have been adopted for use across the OPS:
    - **Date Format** (GO-ITS 74): YYYYMMDD for any date held in government electronic files
    - **Language Code** (Adopted from ISO): 2 digit alpha code (ISO-6390-1) or 3 digit alpha code (ISO-639-2) to identify the language of a Party of interest
    - **Country Code** (Adopted from ISO): 3 digit alpha code (ISO-3166) to represent names of countries
    - **Address Data Standards** (GO-ITS 27.1): These standards specify the data elements and presentation templates that make up Mailing, Delivery and Residential Addresses used in Canada, such as:
        - **Postal Code** - a six-character uniformly structured, alphanumeric code in the form “ANA NAN” where “A” represents an alphabetic character and “N” represents a numeric character. Adopted from Canada Post.
        - **Canada Address** to standardize the data elements and presentation for the civic address (i.e. street addresses), postal box addresses, rural route addresses, and general delivery addresses.
    - The **Common Data Elements Model** (CDEM) developed and supported by the IT Strategy, Policy and Enterprise Architecture Branch of the Ministry of Government Services is a set of reference logical data models on the Party and Address subject areas. The CDEM logical data models are based on ISO and Canada Post Address standards.

- **Ontario Road Network (ORN) Data Standard for Road Geometry and Attributes:** standard dataset that forms part of the foundation of the land information infrastructure for the Province of Ontario. ORN Road Element is currently available for download at no cost from the Land Information Ontario (LIO) Warehouse. Based on ISO 14825:2004.
- **Data Standard for Aboriginal Identifiers:** standard dataset for the application of Aboriginal Identifiers on program and administrative datasets across the OPS. Developed by Ministry of Aboriginal Affairs (MAA).
- **Data Naming Standards:** standardized naming convention for data. GO-ITS 56.3 Information Modeling Handbook (IMH) Chapter 5 provides information about the data naming standards based on ISO / IEC FDIS 11179-1/5. These data naming standards apply to the following data object types:
  - Conceptual and logical data objects that describe the information. Examples are entities, classes, attributes, relationships, operations, and XML elements and attributes.
  - Physical data objects that implement the logical objects. Physical objects should conform to the conceptual and logical standards except where restricted by product limitation. Examples are physical database objects, and XML schema elements.

- **Industry Specific Standards:** These are examples of data standards that have been adopted and are used by a specific OPS business, for instance the health sector within Ontario have adopted national data standards such as:
  - **ICD10** (International Classification of Diseases): a medical classification list developed by the World Health Organization (WHO). It consists of codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases.
  - **SNOMED CT**: a collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting, etc.
  - **HL7** (Health Level 7): An international messaging protocol (format) designed to provide a standard for the exchange of data among healthcare applications.

3.3.3. Data Quality

Data quality is determined by evaluating the ability of the data to fulfill its intended use in the target system. Typically this evaluation is performed by the business to determine

the fit of the data for a particular purpose. For example, the following data quality dimensions taken from the Statistics Canada Quality Assurance Framework (2002) can be used to determine data quality:

- **Relevance** – Refers to how well the data meets the needs of the user. It is concerned with whether the available information sheds light on the issues of most importance to users.

- **Accuracy** – Refers to the degree to which the information correctly describes the subject it was designed to measure.

- **Timeliness** - Refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against accuracy. The timeliness of information will influence its relevance.

- **Accessibility** - Refers to the ease with which the information can be obtained. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.

- **Interpretability** – Refers to the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of data collection and processing, and indications of the accuracy of the statistical information.

- **Coherence** – Refers to the degree to which it can be successfully brought together with other information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology.

3.3.4. Metadata

GO-ITS 46.0 Common Metadata Elements Standard (CMES) defines metadata as **'*data that describes data*** and that enables collaboration and interoperability'. Having metadata makes it easier to find and manage information resources of any type including web documents, databases, films, images, electronic documents and/or paper files. Metadata can contain both business and technical information.

Business metadata describes information in business terms and could include:

- Definitions;
- Business terminology (including business rules) and intended business purpose

- Data ownership; individual or organization that is the authoritative source of the data;
- Data sensitivity;
- Data retention;
- Authorization process/policy for access to data;
- Ranges of valid values for reference data;
- Comments or documentation the data user should be aware of;
- Nature of data; money, date, text, identifier or code, etc., with examples.

Technical metadata describes information in technology terms and could include:

- Data structure; table and field name, data type, maximum length, valid values, business rules, optionality, etc.;
- Indication if the data is raw, calculated, or derived; how it was calculated or derived and the calculation or derivation formula;
- Identifiers; how they are assigned, and by whom;
- Dates; when the data was collected, created, updated;
- Data ownership; name of the individual or organization that is the authoritative source of the data;
- Data sensitivity;
- Data retention;
- Data location; how it can be accessed;
- What validation has been done to the data; what level of data quality the data is;
- Whether ETL/BI transformation logic has been applied;
- Data source; what is the source of the data collected from outside organization(s) (schools, libraries, etc.), collected via surveys of individuals, created in a transactional system, etc.

3.3.5. Analytic Modeling

The term “analytic modeling” usually refers to the process of applying statistics to data (i.e. statistical modeling). The full scope of analytic modeling is much broader and includes statistical modeling for data, problem modeling and solution modeling. Business Analytics includes the creation of statistical models to enable business insights and forecasting.

Statistical models comprise of the following:

- Data description (Input);
- Data transformations required to tailor the data for consumption by the statistical model;
- Statistical model description including the mathematical parameters used;
- Output data description – result set.

A large portion of the effort expended on Analytics is spent on data preparation, which generally consists of the following common steps:

1. Identify the data sets required;
2. Identify the specific data elements from the data sets which are required;
3. Write the code for the first pass which will remove unwanted data elements from the data sets and store the required data elements;
4. Write the code to join the different data sets to create a new data set;
5. Repeat the following steps until the model represents an acceptable working theory and the results provide satisfactory insight;
    a. Run the statistical model with the new data set and compare the results;
    b. Refine the model through additional data transforms;
6. Save the model for future use.

The goal of analytic modeling is to apply the model to derive information and insight from the data for the purpose of:

- Reducing uncertainty in decision processes;
- Identifying and evaluating alternative decisions or actions;
- Reducing complex situations to improve understandability;
- Evaluating risks and rewards of various alternatives;
- Identifying potential consequences and side-effects of decisions and actions.

The model is intended to:

- Structure evidence to organize the facts and reduce complexity;
- Interpret evidence to find meaning hidden in the facts;
- Test hypotheses to seek confirmation or contrary evidence;
- Predict/forecast future events or conditions;
- Simulate what-if scenarios by manipulating variables and observing the results.

Analytic modeling makes extensive use of data, both quantitative and qualitative. Quantitative data uses numbers to express business events, behaviours, and trends as measures (e.g. client count). Qualitative data segments a set of observations (data instances) by categories (e.g. client satisfaction). Analytic data comes from many sources, such as:

- Enterprise Data – Data that is widely used across multiple business functions across the enterprise (e.g. OPS) which is defined, managed, and governed from an enterprise-wide perspective. It usually includes master/reference data, operational data maintained by an IT solution, and the data is contained in a data warehouse and/or a data mart. An example of this would be the IFIS (Integrated Financial Information System) data.

- User Data – Data that is isolated to one or a few business units (e.g. ministry divisions), processes (e.g. serious occurrence reporting process), or functions (e.g. transfer payment management function).
    - Local Data – This is a distinct subset of user data. It may be maintained locally (e.g. stored in Excel spreadsheets), downloaded from a data warehouse and then manipulated to meet a specific request, acquired or derived from external data sources, or generated by earlier analytic processes.
- Industry Data – Data that is from the external data sources such as Statistics Canada or the Canadian Institute for Health Information (CIHI).

Analytics Standards

Predictive Model Markup Language (PMML) is an XML based industry standard developed by the Data Mining Group which has industry support from leading vendors such as SAS, IBM, Oracle and open source R. PMML enables organizations to define statistical models that can be shared among compliant applications.

PMML files include the following main sections:

- **Header**: This area holds basic information about the model such as timestamp of the model's creation and the name of the toolset (software) used to develop the model.
- **Data Dictionary**: This section designates the names of the data variables.
- **Transformations**: This describes any changes made to the original dataset to make it more "modeling-friendly". Sometimes data needs to be fine-tuned (transformed) so that the predictive algorithm can "dive" into the dataset as deeply as possible and extract the most amount of patterns within the data.
- **Model Description**: This is the core of the PMML file. It describes the mathematical parameters of the model. For example, if you export a linear regression model, this section will contain detailed statistical information such as the coefficients used to generate the model's "best fit" curve.
- **Output**: Describes a set of result values that can be returned from a model.

3.3.6. Common Data Elements

Common data elements are a basic set of data elements that are common to most of the business areas within the OPS. The purpose of the common data elements is to enable the information sharing and reuse across the OPS or between ministries with related mandates.

Reference Data

Reference data is data that is defined with a set of permissible values, known as a *value domain,* and used by other data fields. Reference data can be used to classify and categorize other data. Reference data value domains can be defined internally (e.g. Service Desk incident ticket status could be New, Assigned, In Progress, Closed, or Pending) or externally (e.g. industry standards such as the Language Code ISO-6390-1 or Country Code ISO-3166 data standards as referred to in the previous section).

More than one set of reference data value domains may refer to the same conceptual domain. Each value is unique within its own value domain. For example, each province may have:

- An official name (“Ontario”).
- A legal name (“Province of Ontario”).
- A Standard postal code abbreviation (“ON”).
- An International Standards Organization (ISO 3166-2:CA) standard code (“CA-ON”).

Reference data value domains change slowly. Business data stewards should maintain reference data values and associated metadata, and communicate any reference data additions or changes to reference data consumers. Reference data should be a key consideration of data modeling when establishing conformed dimensions (refer to IMH GO-ITS 56.3).

Master Data

Master Data represents the set of data which are agreed on and shared across the enterprise as the single source of basic business data used across multiple systems, applications, and/or processes. Although business rules generally determine the format and allowable ranges of master data values, unlike reference data, master data values are usually not limited to pre-defined domain values. Master data must be the most accurate and authoritative data available to establish the context for data used by key business entities. Common organizational master data includes data about:

- **Parties** such as individuals, organizations, and their roles (e.g. citizens, patients, vendors, employees, students, etc.)
- **Locations** (e.g. addresses)
- **Financial structures** (e.g. general ledger accounts, cost centres, etc. but not transaction specific data)

Master Data Management (MDM) is the process of defining and maintaining how master data will be created, integrated, maintained, and used throughout the enterprise. The three primary areas of focus for MDM are:

1. Identification of duplicate records within and across data sources to build and maintain associated cross-references to enable information integration;

2. Reconciliation across data sources and providing the best version of the truth that enables a merged view of the information across systems to address inconsistencies;
3. Provision of access to "golden" data across applications either through direct reads or by replication feeds to data sources.

Conformed Dimensions

A conformed dimension is a set of data attributes that has exactly the same meaning and content when being referred to from different fact with which it relates. It can be physically referenced in multiple database tables using the same key value to refer to the same structure, attributes, domain values, definitions and concepts. A conformed dimension cuts across many facts.

Dimensions are conformed when they are either exactly the same or one is a perfect subset of the other. If the dimensions are not conformed, different data sets may share what appear to be common codes or identifiers, but which are actually different values and therefore cannot be used to join the data sets and thus much of the potential benefits of the BI and Analytics environment are lost.

A Business Intelligence and Business Analytics program should consider the requirements for conformed dimensions as part of the planning process to ensure that data from multiple sources can be joined and analyzed using the conformed dimensions. Refer to GO-ITS 56.3 (IMH) for more information about conformed dimensions.

3.4. Privacy

Generally speaking, privacy issues are context specific. A Privacy Impact Assessment (PIA) would be necessary in every case involving personal information[11]. Legitimate concerns about the sensitivity of personal data (e.g., clinical health information) require organizations to take appropriate steps to understand the risks and consider methods to depersonalize / anonymize data responsibly such that it reduces privacy risks, protects the privacy interests of individuals and maximizes the value of the data for analytical and research purposes.

There is much debate[12] within the privacy community on whether de-identification can in practice be effective or even possible without the risk of re-identification. While it is important to understand the following concepts within the Anonymization landscape, this reference model does not advocate or support any particular technique or method to depersonalize data responsibly; nor does it suggest that it is possible to depersonalize any data set while maintaining a sufficient level of data quality necessary to be useful. The risk of re-identification is an important privacy impact assessment consideration.

[11] Refer to the Privacy resources on the iNetwork site for further information https://intra.sse.gov.on.ca/inetwork/resourcecentre/Pages/subject.aspx

[12] The following two papers are recommended further reading that argue different sides of the debate: (1) "Big Data and Innovation, Setting the Record Straight: De-identification Does Work", Cavoukian and Castro (2014); and (2) "No silver bullet: De-identification still doesn't work", Narayanan and Felten (2014).

3.4.1. Anonymization

Anonymization is a process that removes the association between the identifying data and the data subject. Many Online Transactional Processing (OLTP) systems contain personal information necessary for the management of their business program. Anonymization provides a means to use longitudinal and linked data required for decision support analytics without compromising the identity of the persons it pertains to.

When anonymizing personal information we need to consider both **direct** and **quasi-identifiers**. Anonymization of personally identifiable information involves removing or modifying either information that identifies an individual (direct identifiers) or related data attributes (quasi-identifiers) that could be used to identify an individual under reasonably foreseeable circumstances, either alone or with other information.

- **Examples of direct identifiers:** Name, address, telephone number, fax number, Medical Record Number (MRN), health card number, health plan beneficiary number, license plate number, email address, photograph, biometrics, Social Security Number (SSN), Social Insurance Number (SIN), device number, clinical trial record number

- **Examples of quasi-identifiers:** sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, total years of schooling, marital status, criminal history, total income, visible minority status, profession, event dates, number of children, high level diagnoses and procedures.

As illustrated in Figure 9, anonymization techniques can generally be divided into two categories: Data Masking for direct identifiers and De-Identification for quasi-identifiers[13]. The choice of method depends on the client business requirements regarding the data, including any need to link the anonymized data back to the individual it pertains to. Where linkage of longitudinal records is required, various methods of anonymization can allow for such linkage without compromising privacy. Anonymization must then be factored into the mitigating controls to reduce the risk of re-identification and a privacy breach.

As an example, anonymization of personal health information (e.g., highly sensitive clinical data), is one mechanism that can reduce privacy risks and increase the availability of the data to analysts. The selection of appropriate anonymization methods requires an understanding of the values, identifiers and attributes for the data, as well as an understanding of how to measure the risk of re-identification of the data, individually and in a collection/linkable datasets. The methods need to consider the purpose for which the data are being provided/exposed, the need to maximize this utility, and the need to minimize the risk of a privacy breach.

[13] Source: presented by Dr. Khaled El Emam, Electronic Health Information Laboratory, CHEO Research Institute

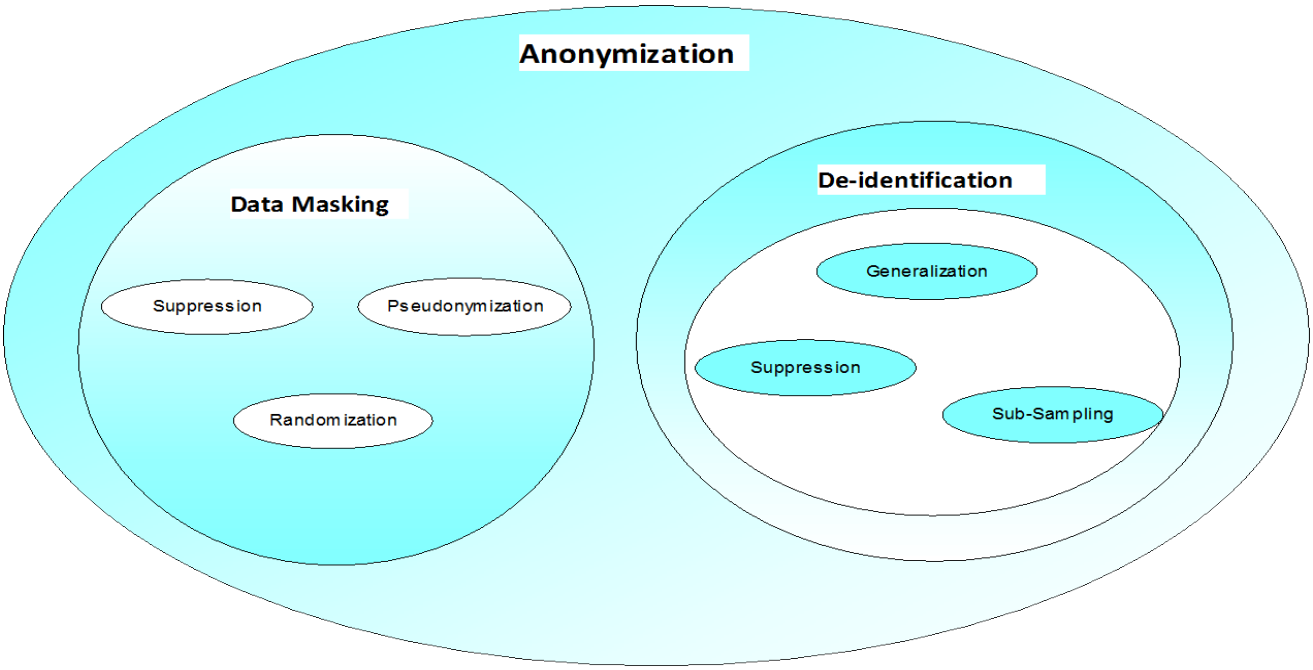

Figure 9: Privacy Architecture

In many cases, it is impossible to completely eliminate all risk of re-identification and still support a BI and Analytics business practice. The aim of anonymization (sometimes referred to as depersonalization of data) within a BI and Analytics practice is to reassure the data steward of an anonymized dataset or repository that there is "no reasonable basis to believe that the information can be used to re-identify the individual" (Guide to the De-Identification of Personal Health Information, Khaled El Emam).

There may be situations where there is a legitimate need to re-identify an individual within an anonymized dataset (e.g., when fraud is detected). In such cases, some anonymization techniques allow for the possibility to reverse the process of depersonalization. The authority and processes to do this needs to be tightly controlled and falls outside the spectrum of the practice of BI and Analytics.

3.4.2. Data Masking

Data Masking involves the application of a set of data transformation techniques to reduce the risk of identifying a data subject to a very small level without any concern for the analytics utility of the data. Data masking obfuscates sensitive data by replacing it with other data; typically characters that will meet the requirements of a system designed to test or still work with the masked results. Data masking ensures that vital parts of personally identifying information are obscured or otherwise de-identified.[14]

[14] Jeff Simpson, IRI Blog. Retrieved 2014-03-11 from URL: http://www.iri.com/blog/data-protection/data-masking-and-data-encryption-are-not-the-same-things/

3.4.2.1. Suppression

Suppression involves the removal of fields from a dataset, thus eliminating the ability to analyze the field values or their relationship to other fields.

3.4.2.2. Pseudonymization

Pseudonymization involves the removal of the identifying association with a data subject by replacing it with one or more artificial identifiers, or pseudonyms. This method can be useful for ensuring the capability to link additional datasets, and datasets that change over time, to be able to conduct longitudinal analyses. This technique allows for the potential to reverse the process of anonymization through a trusted party.

3.4.2.3. Randomization

Randomization involves the replacement of a value in the data with a random value from a large database of possible values.

3.4.3. De-identification

De-identification involves the application of a set of data transformation techniques to reduce the risk of identifying a data subject while retaining a very high level of analytical utility for the data.

3.4.3.1. Generalization

Generalization involves the application of a set of transformation techniques to reduce the precision of a value to a more general one. For example generalization of quasi-identifiers such as date of birth or postal code could involve using just the year and month of birth but not the day, or using just the first three characters rather than the full postal code.

3.4.3.2. Suppression

Suppression involves the removal of identifying records or values in the data is to ensure the anonymity of some individuals without compromising the analytical utility of the data. In situations where even aggregate data would still allow an individual to be identified, the data for those records would be suppressed. For example if the number of female francophone students in a given program was less than five, the related records would be suppressed to prevent the possibility that one of the individuals could be identified.

3.4.3.3. Sub-sampling

Sub-sampling involves randomly selecting a subset of records or individuals from a dataset.

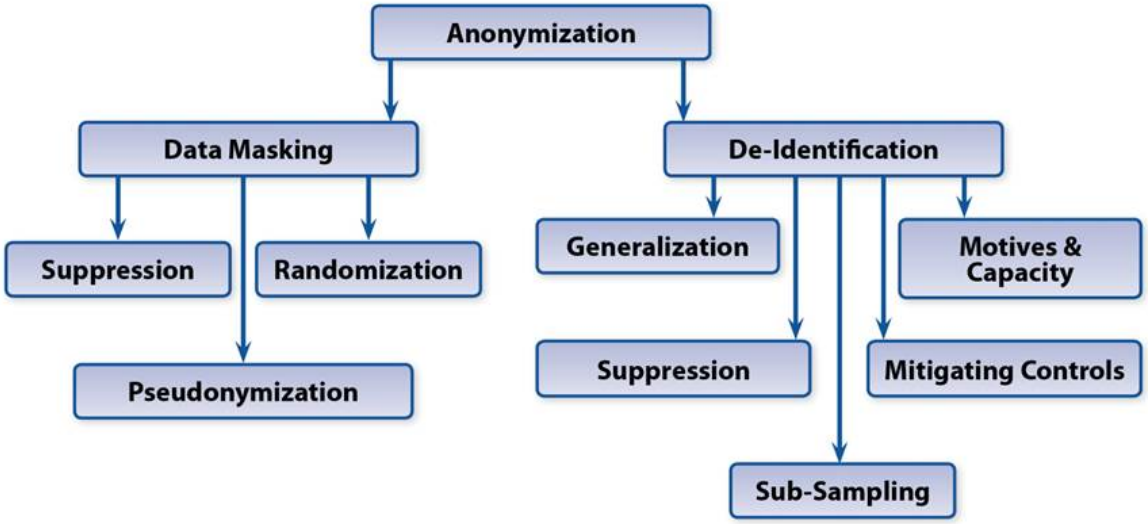

Figure 10: Landscape of Anonymization[15]

As illustrated in Figure 10, within the anonymization landscape the selection and confirmation of de-identification methods includes the consideration of the consumers "**motives and capacity**" to re-identify data and the "**mitigating controls**" that are in place to manage and protect the data which is received. Both these activities could influence the method of de-identification chosen and the risk exposure to re-identification and invasion of privacy.

[15] Anonymization terms and definitions in this section are adapted from Khaled El Emam of Privacy Analytics, Data Anonymization Solutions.

4. Summary

Many organizations in the OPS use or require intelligence in their business to better meet the needs of Ontario citizens. A Business Intelligence and Business Analytics program must consider everything from after the data is recorded to when the information is presented for action or decision-making. The movement to a BI and Analytics business model within an overall data management framework is a complex business transformation.

BI and Analytics solutions across the OPS contribute greatly to business successes by allowing for better and more responsive decision support as well as lower operating costs as a result of a more efficient use of data infrastructure.

The key drivers of this technology include the growing availability of data, and the need for evidence-based decision-making. The increasing amount of digital data available for analysis has the potential to drive innovation, foster greater transparency, and spur economic growth.

BI and Analytics models must be pervasive, comprehensive, and reusable. They are similar in their use of data as measures, use of metrics to evaluate business behaviours, and the shared goals of understanding, insight, and foresight.

BI and Analytics enable organizations to see where their business has been, understand where it is now and predict where it is going.

5. Related Standards

5.1. Impacts to Existing Standards

Identify any GO IT Standards that reference or are referenced by this standard:

GO IT Standard	Impact	Recommended Action
GO-ITS 20.1 *Platform Software Standard*	While reference model is product agnostic, technologies that support BI and Analytics as referenced within this standard may drive GO-ITS 20.1 updates	Maintain alignment between GO-ITS 20.1 and this standard
GO-ITS 30.0 *Database Management Systems*	While reference model is product agnostic, technologies that support BI and Analytics as referenced within this standard may drive GO-ITS 30.0 updates	Maintain alignment between GO-ITS 30.0 and this standard
GO-ITS 30.3 *OPS Business Intelligence Software*	While reference model is product agnostic, lead organization / standard owner and stakeholders are to provide lessons learned vis-à-vis gap-fit analysis between reference model and product standard – important for refinement of the model over time. Updates to this reference model may drive GO-ITS 30.3 updates.	Maintain alignment between GO-ITS 30.3 and this standard

6. Compliance Requirements

In order to manage the effectiveness of this standard, then ministries, clusters, and applicable agencies are expected to adopt and monitor compliance.

7. Roles and Responsibilities

Accountable Role (Standard Owner) Definition
The individual or committee ultimately accountable for the process of developing and maintaining this standard. Where a committee owns the standard, the committee Chair is accountable for developing the standard including future updates. There must be exactly one accountable role identified.

Accountable Role:
Title: Head, Data Collection & Decision Support Solutions (DDSB)
Ministry/I&IT Cluster: Community Services I&IT Cluster (CSC)

Responsible Role Definition
The organization(s) responsible for the development of this standard. There may be more than one responsible organization identified if it is a partnership/joint effort. (Note: the responsible organization(s) provides the resource(s) to develop the standard).

Responsible Organization(s):
Community Services I&IT Cluster (CSC), Data Collection & Decision Support Solutions
Labour and Transportation I&IT Cluster (LTC), Road User Safety Solutions
Health Services I&IT Cluster (HSC), I&IT Strategy and Architecture Branch

Support Role Definition
The support role is the resource(s) to whom the responsibility for developing and maintaining this standard has been assigned.

Support Role (Editor):
Ministry/I&IT Cluster: CSC
Branch: Data Collection & Decision Support Solutions
Section: Enterprise Business Intelligence Unit
Job Title: Senior Business Performance Specialist
Name: Binh Lu
Phone: 416-325-2300
Email: Binh.Lu@ontario.ca

Working Group Members:
Stephen Milling (HSC), Ajay Gupta (HSC), Irene Gronek (HSC), Louise Doyon (HSC), Danna Dobson (HSC), Lori Burns (HSC), Martin Hodgkinson (HSC), Joanne Venema (HSC), Sonya Skinner (CSC), Binh Lu (CSC), Andrew Sally (CSC), Sheila Gu (CSC), Lucy Cao (CSC), Suzanne Bond (CSC), Tim Leblanc (ITS), Laura Pastirik (ITS), Tariq Hasan (LTC), John Lee (CYSSC), Daniel Chang (CYSSC), Ling Chen (CYSSC), Marianne Sutcliffe (JTS), Louis Mick (JTS), Thomas Chen (MGCS), Dean Pigeon (SCS), Paul Daly (SCS), Linda Jarvis (SCS), George Marzotto (OSS), Paul Bennett (CAC), Lisa Poirier (CAC), Nancy Chen (SO)

8. Consultations

Areas consulted as part of the development of this standard include individuals and committees, councils and/or working groups:

Organization Consulted	Division	Branch	Date(s)
MCSS/MCYS/CYSSC	Strategic Policy and Planning; Social Policy Development	Strategic Information and BI; Systems Development and Maintenance; Social Services Solutions Modernization; Policy Research and Analysis	2013-03-04 2013-10-01 2013-10-21 2013-10-23
MAG/MCSCS/JTS	Corporate Services Management;	Controllership and Reporting; Business and Fiscal Planning; Common Cluster Solutions	2013-02-20 2013-10-10 2013-10-28
MTO/LTC	Data Mgmt & Operations	Road User Safety Solutions; Architecture, Information Mgmt & Labour Solutions	2013-03-08
MNR/LRC	Corporate Management and Information	Data Management and Reporting; Mapping and Information Resources	2013-02-21 2013-10-04 2013-11-19
MOHLTC/HSC	Health System Information Mgmt and Investment	I&IT Strategy and Architecture; Integrated Health Solutions; Health Data Branch	2013-01-10 2013-08-14
MTCU/MEDU/CSC	Strategic Policy & Programs; Employment & Training; Student Achievement	Data Collection & Decision Support Solutions; CGMS; Research & Planning; Literacy & Numeracy Secretariat; Service Standards & Accountability	2013-09-30 2013-10-04
MGS/EFSS/CAC/ITS SCS/IPA	Ontario Shared Services; Enterprise Financial Services & Systems	Customer Solutions Delivery; Strategic Partnerships & Program Policy; IFIS BI and Reporting; Portfolio & Project Management Office; Open Government Project	2013-02-20 2013-02-22 2013-02-27 2013-10-07
ServiceOntario/SOIT	Business Improvement	Business Effectiveness; Business Solutions	2013-10-02 2013-10-30

Committee/Working Group Consulted	Date(s)
Solutions Delivery Leadership Council (SDLC)	2013-11-21 and 2014-08-11
Information Management Leadership Committee (IMLC)	2013-10-02 and 2014-04-09
Information Architecture Domain Working Group (IADWG)	2013-09-17 to 2014-04-15
Business Architecture Domain Working Group (BADWG)	2013-09-19 to 2014-07-03
Corporate Architecture Core Team (ACT)	2013-11-20 and 2014-07-30

9. Appendices – Related Business Architecture Examples

9.1. Resource Type

Resource Type lists and defines the types of resources that are important to the business that are either consumed by some business processes or produced by them.

Resource Type	Sub-Type	Resource	Description
Material	Software Tool	BI Tools (Business Intelligence Software)	Software acquired from a vendor to provide business intelligence capabilities to clients.
		Analytics Tools (Business Analytics Software)	Software acquired from a vendor to provide business analytic capabilities to clients.
		Extract Transform Load (ETL) Tools	Software tools that are used for exporting, transforming and loading data.
		Database Management System (DBMS) to run Data Warehouse/Data Mart	Software that organizes, stores, and retrieves data from various sources required to provide BI and Analytics capabilities.
		Agile Search and Discovery Tools	Software that enables the client to combine unstructured and structured data from multiple sources and perform analysis on the raw data to yield new information and insights regarding the business.
	Licence	Software/Hardware Licences	Governs the usage or redistribution of software and hardware.
Information	Requirement	Business Requirements	Details what must be delivered (by way of new or revised process(es), software or systems) in order to provide desired business value.
		Information Requirements	Details the information that is required to achieve desired change in business, as expressed in the business requirements.

Resource Type	Sub-Type	Resource	Description
	Data	Source Data	Data from internal or external sources that is submitted and loaded to the repository.
		Production Data	Data that has been transformed according to predefined quality and data integration processes.
		Conformed Dimension Data	The data content of the data attributes that serve as a conformed dimension.
		Metadata	Data about data; the metadata content specifies the source of the dataset, the nature of the dataset uses; accuracy of the dataset.
	Business Intelligence Outputs	Pre-defined Report	Reports that are generated by the tool on a predefined schedule and format.
		Dashboard	A real-time user interface, showing a graphical presentation of the current status (snapshot) and historical trends of an program's key performance indicators (KPIs) to enable instantaneous and informed decisions to be made at a glance.
		Ad-hoc Query	Reports that are customized based upon client needs and generated by the tool as per client's timeline.
		Scorecard	A graphical representation of the progress over time of some entity, such as a program or a business unit, toward some specified goal or goals.
		Graphics	A graphic display that enables the user to visualize, question, analyze, interpret, and understand data to reveal relationships, patterns, and trends.
		GIS Integration	Reporting that shows data from a geographical location perspective.

Resource Type	Sub-Type	Resource	Description
		Key Performance Indicator (KPI)	A business metric used to evaluate factors that are crucial to the success of an organization. KPIs are applied in business intelligence (BI) to gauge business trends and advise tactical courses of action.
	Business Analytic Outputs	Regression Model	A statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.
		Longitudinal Analysis	A study of a population over time.
		Econometric	The application of statistical and mathematical methods in the field of economics to describe the numerical relationships between key economic forces such as capital, interest rates, and labor.
	Agile Search & Discovery Outputs	Structured and Unstructured Data Analysis	A study of structured and unstructured data in the pursuit of agile insights.
		Interactive Visualization	A visual representation of data which enables clients to add or modify linkages amongst data to uncover new knowledge.
		Sentiment Analysis	Use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.
	Data Model	Conceptual Data Model	A formal representation of the data needed to support business processes and functions delivered by an enterprise or a business domain. Includes a diagram and supporting data dictionary.

Resource Type	Sub-Type	Resource	Description
		Fact and Dimension Matrix	A high-level view of business information requirements that are in scope for a decision support solution. The matrix illustrates different data analysis perspectives (i.e., dimensions). It is used to document and communicate the scope of the data warehouse and/or data marts supporting a decision support solution. Includes a matrix and supporting descriptions.
		Logical Dimensional Model	A logical level data model that includes the details of the fact entities, facts, dimension entities, and the relationships connecting the fact entities within the dimension entities. Includes a diagram and supporting data dictionary.
		Physical Dimensional Model	Description of the internal data structures used by the data mart. It is used to define the physical database objects for implementation specific to the selected RDBMS Includes a diagram and associated data dictionary.
		Data Warehouse Logical Model	A model that represents the logical data structures to capture non-volatile data, in scope snapshot entities, their relationships, their primary attributes and some calculated attributes. Includes a diagram and a data dictionary.
		Data Warehouse Physical Model	A model that describes the internal data structures, used by the data warehouse. Includes a diagram and a data dictionary.

Resource Type	Sub-Type	Resource	Description
	Document	Business Scenarios	A description of the expected and alternate sequences of events. It is used to test and elaborate designs and is the basis for fleshing out and defining business requirements.
		Training Material	Materials that support the delivery of a training course.
		Design Document	Documents detailing design specifications for delivering the decision support service.
		Interface Data Requirements Document	Documents detailing exchanges of information between systems, and showing data lineage between source and target.
	Agreement	Data Sharing Agreement	Agreement that sets out the terms and conditions for the use of data by or from third parties.
		Service Level Agreement	Describes the services and performance standards agreed upon between OPS and the vendors, or between organizations within the OPS, providing products (software, hardware) and services
		Memorandum of Understanding	Sets out the terms and conditions for the provision of information management and information technology services between Service Owner and client.
	Request	Access Request	A request made by a client for access to a specific Business Intelligence or Business Analytics service.
		Training Request	A request for training to be delivered.
		Request for Assistance	A request issued by a client for assistance with a Business Intelligence or Business Analytics service.
		Escalated Incident Request	A request for assistance issued by a client that has been transferred to another tier in the support model for resolution.

Resource Type	Sub-Type	Resource	Description
	Response	Access to Information Permit	A period of permission to access data granted to a client.
		Delivery of Training	The delivery of training with respect to a specific Business Intelligence or Business Analytics service or capability.
		Resolved Issue	A response to a request to assistance which rectifies the problem experienced by the client.
	Schedule	Data Submission Schedule	The schedule that determines when (and how often) source data is to be submitted for loading into the Data Warehouse.
	Specifications	Analytics Design Specifications	Specifications for reports produced from the repository.
	Training Materials	Course Material	Materials that support training course delivery.

9.2. Party Type

Party Type identifies and classifies parties of interest to the service, to help ensure that all party types are accounted for when conducting needs analysis for a given ('As Is' or 'To be') service. The parties listed below provide examples of individuals and organizations that are relevant to the notional BI and Analytics Program. Not all parties listed will necessarily be applicable and additional parties may be required to support unique program requirements.

Party Type	Subtype	Party	Description
Organization			A group of individuals having a specific purpose.
	OPS Organization		An Organization that is part of the Ontario Public Service (OPS), meaning that it employs public servants under the Public Service Act of Ontario, which provides public programs and services in support of the public's needs for the province of Ontario.

Party Type	Subtype	Party	Description
		Ontario Ministry	An organization through which a political unit exercises authority and performs functions and which is usually classified according to the distribution of power within it. Examples include: Ministry of Health and Long-Term Care, Ministry of Government Services.
		Ontario Ministry Division	An organization unit of the Province of Ontario that supports a Ministry's mandate. Examples include: Health System Information Management and Investment Division of the Ministry of Health and Long-Term Care, Infrastructure Technology Services of the Ministry of Government Services.
		Ontario Ministry Branch	An organization unit of the Province of Ontario that supports a Division's mandate. Examples include: Health Data Branch and Health Analytics Branch of the Health System Information Management and Investment Division.
		Ontario Ministry Unit	An organization unit of the Province of Ontario that supports a Branch’s mandate.
	Other OPS Organization – Non-Ministry		
		I&IT Cluster	An organization that supports the I&IT requirements of its client Ministries.
	Broader Public Sector Organization		An organization that has an arms-length relationship with the government. That is, it receives public funds, created by specific legislations, but is not part of the government. Examples include: Colleges and Universities, Local Health Integration Network (LHIN), Hospitals, etc.

Party Type	Subtype	Party	Description
	Non-Government Organization	Resource Provider (e.g., vendor)	An organization that provides a service or item for a set price or fee structure. Examples: Third-Party non-OPS tool, hosting or value provider (e.g. myBlueprint, SAS, IBM, etc.).
		Information and Privacy Commissioner Ontario (IPC)	An independent oversight body responsible for ensuring compliance with the Province's access and privacy legislation (FIPPA, MFIPPA, PHIPA) and upholding and promoting open government and the protection of personal privacy. As part of its mandate, the IPC comments on proposed government legislation and programs, and investigates privacy breaches.
		Regulatory Body	A body that regulates a profession by establishing and enforcing standards of practice. Examples include: College of Nurses of Ontario, College of Physicians and Surgeons of Ontario (CPSO).
		Professional Association	A group of individuals that has a common interest to support the individuals they represent and to promote their area of expertise. Examples include: Ontario Medical Association (OMA); Ontario Nurses Association (ONA).
		Research Partner	An organization that aims to contribute to the improvement of the same target group as the Program. E.g. Canadian Institute of Health Information (CIHI) has a shared interest in analyzing health records and is reliant on quality information to conduct their research.
Individuals		Ontario Resident	An individual who resides within the Province of Ontario.
		Non-Ontario Resident	An individual who resides outside of the Province of Ontario

Party Type	Subtype	Party	Description
		OPS Staff	An individual who works for an OPS Organization and who provides services to achieve the goals of their program.
		Non-OPS Staff	An individual who does not work for an OPS Organization but who needs to interact in some way with the program.

9.3. Role Type

Role Type lists the types of roles played by parties (individuals and organization) of interest to the business. A role is defined by a set of functions or relationships played by a party. The following are additional examples of roles that may be relevant to the business processes executed as part of program and service delivery for the notional BI and Analytics program in the OPS.

Further analysis with respect to a project's specific scope is required in order to rationalize the Roles suggested below.

Role	Role Sub-Type	Description	Party or Parties playing the role
Business Intelligence and Analytics Designer		An individual that specializes in planning, designing, building and implemented Business Intelligence and Analytics tools and supporting frameworks.	E.g., OPS Staff
Data Administrator		An individual who has been given the responsibility of managing the data but is not the owner or creator of the data.	E.g., OPS Staff
Data Analyst		A subject matter expert who analyzes and abstracts data and circulates results (datasets and reports) to information clients.	E.g., OPS Staff
Data Collector		An individual or organization that receives data from the Data Contributor, verifies and validates the quality of the data before it is saved to the repository.	E.g., Ontario Ministry Division
Data Consumer		An individual or organization within or outside of the enterprise that obtains data from the BI and Analytics program.	
	Internal Data Consumer	An individual or organization within the OPS that consumes data obtained from the BI and Analytics program.	E.g., OPS Staff

Role	Role Sub-Type	Description	Party or Parties playing the role
	External Data Consumer	An individual or organization outside of the OPS that consumes data obtained from the BI and Analytics program.	E.g., Non-OPS Staff, Other OPS Organization - Agency
Data Contributor		An individual or organization that contributes data to the BI and Analytics program.	
	Internal Data Contributor	An individual or organization within the OPS that provides existing data used to support their program to the BI and Analytics Program.	E.g., Ontario Ministry Branch
	External Data Contributor	An individual or organization outside of the OPS that provides data to the BI and Analytics Program.	E.g., Other OPS Organization - Agency
Data Custodian		An individual or department that is responsible for the safekeeping, transport, data storage and implementation of business rules, based on the applicable Data Sharing Agreements, instructions/decisions received from the Data Steward/governance body, and in line with direction provided by the GO-ITS standards or other OPS and Ontario rules/legislation.	E.g., Ontario Ministry Division, Ontario Ministry Branch
Data Owner		An individual or department that has the accountability for the data generated and/or consumed. Data owners are generally people who initially created or produced the data, and/or have authority and accountability under legislation, regulation, policy or other instrument for that particular data. The Data Owners make decisions on the data such as whether to provide data for disclosed uses, what safeguards should be in place to manage data risks, classification and access decisions regarding the collection, transformation, use, retention and disposal.	
Data Planner		An individual who plans the acquisition of data.	E.g., OPS Staff

Role	Role Sub-Type	Description	Party or Parties playing the role
Data Quality Assurance Analyst		An individual who identifies and resolves data gaps, redundancies and other data quality issues.	E.g., OPS Staff
Data Steward		An individual or department that serves as the appointed trustee for data assets. Data Steward has the delegated responsibility to manage the Owner's data and its usage on the Owner's behalf. Where the Data Steward is not also the Data Owner, the Data Steward is bound by Data Sharing Agreement and/or other agreements and disclosures made with the Data Owner when the data was collected for use.	
Decision Support Analyst		An individual that specializes in designing and analyzing decision support enablers (e.g. Business Analytics) and the infrastructure required to support them.	E.g., OPS Staff
ETL Specialist		An individual that specializes in Extra, Transform and Load procedures and best practices.	E.g., OPS Staff
Information Consumer		An individual or organization within or outside of the enterprise that obtains information from the BI and Analytics program.	
	Internal Information Consumer	An individual or organization within the OPS that obtains information from the BI and Analytics program.	E.g., OPS Staff
	External Information Consumer	An individual or organization outside of the OPS that obtains information from the BI and Analytics program.	E.g., Non-OPS Staff, Other OPS Organization - Agency
Ontario Taxpayer		A resident of Ontario that pays taxes to the government to support its portfolio of programs and services.	E.g. Ontario Resident
Policy Analyst		An individual engaged in setting information policies and procedures.	E.g., OPS Staff
Privacy Analyst		An individual, who manages privacy issues, enforces and performs assessments in compliance with privacy standards.	E.g., OPS Staff

Role	Role Sub-Type	Description	Party or Parties playing the role
Program Manager		An individual who manages program's objectives and enforces data governance across the program.	E.g., OPS Staff
Program Owner		A party accountable for achieving the goals of the program.	E.g., Ontario Ministry Division
Security Administrator		An individual who ensures access data and/or information resources is granted to authorized individuals and that access audits can be performed.	E.g., OPS Staff
Security Analyst		An individual who manages security issues, and performs assessments in compliance with security standards.	E.g., OPS Staff
Service Owner		An individual or organization that is accountable for the effective and efficient delivery of the service to clients (e.g., owner of tools and capabilities for BI and Analytics).	E.g., Ontario Ministry Division
Service Provider		An individual or organization that is accountable for the effective and efficient delivery of services to clients and the resolution of service incidents (e.g., provider of BI and Analytics tools and capabilities to clients).	E.g., Ontario Ministry Branch
Tools Analyst		A subject matter expert who provides expertise in the setup, use and dissemination of tools used in the Program.	E.g., OPS Staff

9.4. Target Group Type

A target group type is a classification of that part of the population whose needs the program has a mandate to satisfy. By classifying target groups, program owners can make explicit decision about how their needs will be met, including service delivery channels and security and privacy requirements. To ensure the business architecture for the notional BI and Analytics program can be leveraged and extended by Business Architecture practitioners, determining target group types stemmed from analysis of the needs satisfied by services owned by the program area.

In general, target groups can exist within (internal) or outside (external) the enterprise.

Target Group Type: Client Group	Description

Data Consumer	An individual or organization within or outside the enterprise that obtains data from the BI and Analytics program.
Information Consumer	An individual or organization within or outside the enterprise that obtains information from the BI and Analytics program.
Target Group Type: Interested Party	**Description**
Individuals (E.g. Ontario Taxpayer)	A person who is indirectly affected by the program's service offerings.

9.5. Need Type

Need Type identifies the needs of a target group that the program intends to satisfy. It is expected that this list of needs will be used as a starting point for any initiative that involves BI and Analytics and extended to further describe existing needs or include needs not yet defined.

In our notional program, the Data Provision Service is owned by the BI and Analytics Program. The accountability for the delivery of this was included in the notional program as it implies a tight coupling between data management practices and standards and the ability to collect, integrate, and store within a repository (e.g. Data Warehouse and Data Marts) information that can further be extended to yield new knowledge in a variety of formats.

Need	Description
Accurate Data	Data that is of the required quality to support the business information requirements.
Actionable Conclusions	Information and analysis (conclusions) about the current or potential future state of an enterprise that can be used to alter aspects of the business (e.g., business processes, service delivery) to achieve intended goals.
Access to Quality Information	Access to quality information to support and improve business performance.
Agile Search and Discovery Mechanism	Tools or processes that provide the business with the ability to view and analyze data.
Availability of Business Intelligence and/or Analytics tools and support	Access to specialized tools that permit the discovery of new information or knowledge.
Choice	Refers to the degree of choice presented for tailoring the output and its delivery. Examples of choice include delivery channel, specification parameters.[16]
Reliable Information	Access to information that will permit the analysis as to a program's effectiveness or efficiency with respect to

[16] Adapted from the GSRM Service Performance Metrics, BTEP.

Need	Description
	service delivery and target group need satisfaction.
Risk Awareness	Awareness of the risks associated with enterprise assets. E.g. Risk of disclosing personal information.
Risk Mitigation Strategies	Strategies for mitigating risks identified to enterprise assets.
Timely Information	Information that is available when required to ensure decision makers have ample opportunity to consider all courses of action based upon receipt of this new knowledge.
Valuable Insights	Information that when analyzed provides the business with a new perspective or knowledge. When actioned, these insights can lead to improvements in the program, service or process efficiencies and/or effectiveness.

9.6. Goal

Goal expresses a desired change to a target group (e.g. client group). Program goals state the desired change to a target group and are expressed as outcomes (measureable results directly attributed to the program and impacts (results influenced by the program)[17].

Goal(s):	Outcome(s)	Impact(s)
Support evidence based decision-making	Identification of Customer centric patterns and Operational patterns	Improved understanding of target group for more effective actions. Improved budget forecasting and operational efficiency.
Skilled use of common methodologies and data standards	Consistent analytical methodology based on best practices.	Increased client satisfaction.
Provide clients with more self-service capabilities	Minimizes the reliance of business on Information Technology to provide information products to meet their needs.	Increased client satisfaction.
Strategic capture, management, protection and use of data and information	Data is current, accurate and adheres to standards and policies.	Improved decision-making at both the client and program levels.
	Data is well managed / shared / accessible and complies with policies and standards.	Improved reliability, completeness and trustworthiness of knowledge derived from the program.

[17] GO-ITS 56 Appendix B

9.7. Strategies

Strategy identifies how the program will achieve its mandate, reduce or manage risk and achieve desired outcomes[18].

Program Management Strategy	
Knowledge Management	Optimize the creation, capture, storage, maintenance, retrieval and protection of key business knowledge to support both operational activities and performance monitoring.
Data Resource Management	Manage program resources (e.g., data) through the establishment of practices and processes such as Data Governance practices, Data Standards enforcement and Data Life Cycle Management processes.
Program Delivery Strategy	
BI and Analytics Collaboration	Provide the capability to support holistic, cross-program support for both the business processes performed by, and the information required by clients requiring actionable conclusions from BI and Analytics initiatives.

9.8. Business Rule Source

Business Rule Source is used to document the authoritative sources of Business Rules and provide information about those sources[19]. The following list comprises common sources for business rules that constrain business processes that form the service delivery chain. Business Rules for BI and Analytics are largely based on legislated business policies and procedures that control how data and information are created, shared, and disclosed.

For example, a program may have a business rule on how to identify what qualifies as an English language learner using a number of factors such as country of origin.

Business Rule Source		
Source Type	**Name**	**Description**
Legislation	Freedom of Information and Protection of Privacy Act (FOIPPA)	Sets out the legal requirements respecting access and privacy for ministries and agencies.
	Personal Health Information Protection Act (PHIPA)	Establishes a comprehensive set of rules about the manner in which personal health information may be collected, used, or disclosed across Ontario's health care system.

[18] GO-ITS 56 Appendix B
[19] GO-ITS 56 Appendix B

Business Rule Source		
Source Type	**Name**	**Description**
	Archives and Recordkeeping Act (ARA), 2006	Ensures that the public records of Ontario are managed, kept and preserved in a useable form for the benefit of present and future generations.
	Personal Information Protection and Electronic Documents Act 2000 (PIPEDA)	Governs how private sector organizations collect, use and disclose personal information required for program delivery (federal legislation).
	Accessibility for Ontarians with Disabilities Act 2005 (AODA)	Requires all organizations, both public and private, that provide goods or services either directly to the public or to other organizations in Ontario (third parties) and that have one or more employees to provide accessible customer service to persons of all ability levels.
Policy	Corporate Policy on Recordkeeping	Defines and establishes the organization's responsibilities for the creation and management of business records.
	Corporate Policy on the Protection of Personal Information	Establishes the policy and control parameters for the management and use of information and information technology (I&IT) in the OPS.
	Information Security and Privacy Classification Policy (ISPC)	Identifies the requirement for the classification of all information according to their sensitivity – unclassified, low, medium or high
Agreement	Data Sharing Agreement	Agreement that sets out the terms and conditions for the use of data by or from third parties.
	Service Level Agreement	Describes the services and performance standards agreed upon between OPS and the vendors, or between organizations within the OPS, providing products (software, hardware) and services.
	Memorandum of Understanding	Sets out the terms and conditions for the provision of information management and information technology services between Service Owner and client.

10. Glossary

Term	Definition
Anonymization	A process that removes the association between the identifying data and the data subject. Can also be referred to as the depersonalization of data.
Big Data	High volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. The collection of datasets so large and complex that it becomes difficult to process using traditional data processing applications. Source: Gartner
Business Intelligence	The transformation of raw data into meaningful and useful information used to enable more effective strategic, tactical and operational insights and decision-making that contributes to improving overall enterprise performance. Source: Adapted from Forrester, “Craft Your Future State BI Reference Architecture”. November 2012. Evelson and Yuhanna.
Business Analytics	The identification of meaningful patterns and correlations among variables in complex, structured and unstructured, historical, and potential future datasets for the purposes of predicting future events and assessing the attractiveness of various courses of action with the goal of highlighting useful information, suggesting conclusions and supporting decision-making. Source: Updated definition based on Gartner, Forrester and GO-ITS 56.4 BI Reference Model version 1.0, 2009
Conformed Dimensions	A set of data attributes that have been physically referenced in multiple database tables using the same key value to refer to the same structure, attributes, domain values, definitions and concepts. A conformed dimension cuts across many facts. Source: Ralph Kimball, Margy Ross, The Data Warehouse Toolkit (2002)
Data	A collection of facts, concepts or instructions formatted in a special way suitable for communication, interpretation, or processing by human or automatic means. It defines the type of information exchanged between business processes Source: OPS Enterprise Architecture Glossary

Term	Definition
Data Collection	Refers to the formal and highly structured set of data related services that allows organizations to define, manage, collect, validate and report on a single, unified and trusted view of quality data for the purpose of policy formation, program evaluation and performance assessment.
Data Governance	The exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. Data Governance is high-level, executive data stewardship, focused exclusively on the management of data assets. Data Governance maximizes the value of data assets by defining/controlling what enterprise wide level decisions need to be made regarding corporate data, how decisions are made, and what roles are involved to coordinate the demands and interests of multiple stakeholders across the enterprise. Source: Data Management Body of Knowledge (DMBOK)
Data Integrity	A measure of the consistency and accuracy of data. The maintenance of rules that prevent accidental or deliberate alternations or destructions of the data content in a database that would affect its accuracy or reliability. Source: OPS Enterprise Architecture Glossary
Data Management	The practices, architectural techniques and tools for achieving consistent access to and delivery of data across the spectrum of data subject areas and data structure types in the enterprise, to meet the data consumption requirements of all applications and business processes. Source: Gartner
Data Quality	Data Quality is the degree to which data is fit for its intended purpose, measured across multiple dimensions such as relevance, accuracy, timeliness, and completeness. By definition, data quality is subjective based on the relevance of information to users' needs, and characteristics of the information such as accuracy, timeliness, and completeness that affect how it can be used. Source: Statistics Canada Quality Assurance Framework (2002)[20]

[20] Statistics Canada Quality Assurance Framework (2002) defines six dimensions of quality: relevance, accuracy, timeliness, accessibility, interpretability, and coherence. The Canadian Institute for Health Information Data Quality Framework (2009) defines five dimensions: relevance, accuracy, timeliness, comparability, and usability.

Term	Definition
Data Warehouse	A database, or collection of databases, which serve as a central repository for all or significant parts of the data that an enterprise's various business systems collect. Data warehousing emphasizes the capture of data from diverse sources for useful analysis and access, but does not generally start from the point-of-view of the end user or knowledge worker who may need access to specialized, sometimes local databases. The latter idea is known as the data mart.
De-identification	A method of anonymization that involves the application of a set of data transformation techniques to reduce the risk of identifying a data subject while retaining a very high level of analytical utility for the data.
Extract, Transform, Load (ETL)	In managing databases, extract, transform, load (ETL) refers to three separate functions: • First, the extract function reads data from a specified source database and extracts a desired subset of data. • Next, the transform function works with the acquired data - using rules or lookup tables, or creating combinations with other data - to convert it to the desired state. • Finally, the load function is used to write the resulting data (either all of the subset or just the changes) to a target database, which may or may not previously exist Source: Search Data Management (per IADWG)
Information Management	Applies common management principles (planning, directing, controlling, evaluating) to information assets in order to support value-added work. It means establishing disciplined and consistent practices related to the end-to-end management of information throughout its lifecycle: planning, collection or creation, use and dissemination, maintenance, disposition, and evaluation. Source: IPA
Master Data	Master Data represents the set of data which are agreed on and shared across the enterprise as the single source of basic business data used across multiple systems, applications, and/or processes. Also known as the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts. Source: Gartner

Term	Definition
Master Data Management	A technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official shared master data assets.
Open Data	A concept in which low-risk, non-confidential government data is made available to the public so that citizens and businesses can repurpose the data to develop new ideas, services and applications. For data to be "open data" it needs to be made available at no cost, in a digital format and under a flexible use license that carries few conditions. Source: OPS Open Data Publishing Guidebook (2013); for more information refer to www.ontario.ca/opendata
Pseudonymization	A method of anonymization that involves the removal of identifying associations with a data subject by replacing it with one or more artificial identifiers, or pseudonyms. Can also be referred to as tokenization (substituting a sensitive data element with a non-sensitive equivalent, referred to as a token, which has no extrinsic or exploitable meaning or value).
Risk Management	A systematic approach to setting the best course of action under uncertainty by identifying, assessing, understanding, acting on, monitoring, and communicating risk issues. Integrated risk management is a continuous, proactive and systematic process to understand, manage and communicate risk from an organization-wide perspective. It is about making strategic decisions that contribute to the achievement of an organization's overall corporate objectives. Source: OPS Risk Management Policy (MOF)
Reference Model	An abstract framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. Source: Organization for the Advancement of Structured Information Standards (OASIS) (www.oasis-open.org)
Sentiment analysis	Also known as opinion mining, refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.[21]

[21] Definition of sentiment analysis sourced from: http://en.wikipedia.org/wiki/Sentiment_analysis