City of Philadelphia

Department of Behavioral Health and Intellectual disAbility Services

DATA GOVERNANCE FRAMEWORK STRATEGIC PLAN

VERSION 2.03

18 DECEMBER 2018

Guidelines for Aligning Information Management Concepts, Practice and Context

DBHIDS

DEPARTMENT of BEHAVIORAL HEALTH and INTELLECTUAL disABILITY SERVICES

DAVID T. JONES

COMMISSIONER

TABLE OF CONTENTS

Table of Contents 2
Executive Summary 3
Preamble to the 2017 Update 4
The Need for a Data Governance Framework 7
What is a Data Governance Framework? 7
Data Governance Framework as Information Architecture 8
Data Governance 1.0 versus Data Governance 2.0 15
Transition to Data Governance Framework Version 2 16
The Data Management Landscape Prior to 2015 16
Status of DBHIDS Data Governance Strategies, Version 1 17
The Data Management Landscape in 2017 19
Strategic Plan Components 20
The DBHIDS Data Governance Framework Strategic Plan 22
DBHIDS Data Governance Mission 22
DBHIDS Data Governance Vision 22
Data Strategy Formulation 22
DBHIDS Information Architecture Principles 24
DBHIDS Data Governance Goals 26
Data Management Strategies 28
DBHIDS Data Architecture Levers 29
DBHIDS Information Service Delivery Use Cases 29
DBHIDS Conceptual Information Architecture Schema 30
Moving Forward 31
Works Cited 33
Credits 35
Appendix A: Glossary of Information Architecture Terms 36
Appendix B Ladley's Generally Accepted Information Principles ™ 58
Appendix C: Strengths, Weaknesses, Opportunities, and Challenges (SWOC) Analysis 59
Appendix D: Data Strategy Assessment 65
Appendix E: Data Governance Roles and Responsibilities 66
Appendix F: The Owner/Steward/Custodian Model 71

DBHIDS © 2015-2019 – All Rights Reserved

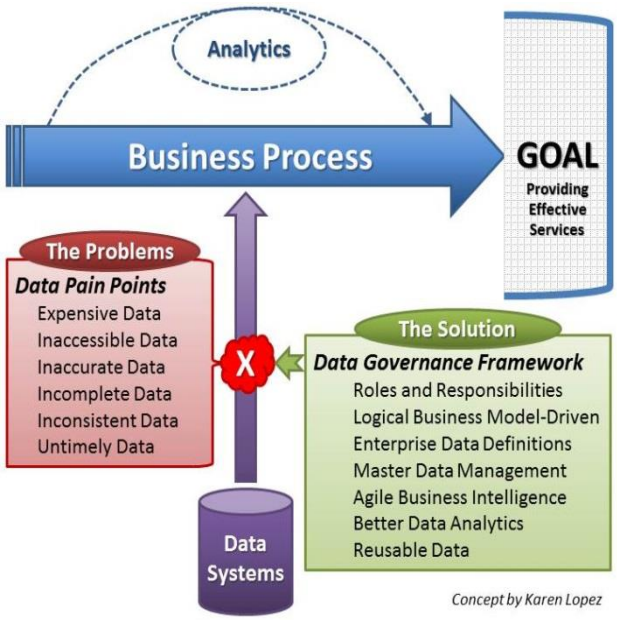
Daniel J. Paolini, CIO – with input from the Enterprise Data Management Team

"We are drowning in information but starved for knowledge. Uncontrolled and unorganized information is no longer a resource in an information society, instead it becomes the enemy."

— John Naisbitt, *Megatrends: 10 New Directions Transforming Our Lives* (1982)

EXECUTIVE SUMMARY

Information Technology is the Information, Applications and Infrastructure necessary to meet the needs of the Business. Of these, the most important resource is the information asset. It is this information that represents a model of the functioning business over time. Infrastructure and applications change to reflect changes in technology or process, but the data that records the transactions of key entities with the business must be logically defined and maintained so that it is consistent, persistent and useful.



Data quality degrades when data is not managed continually by the business. The ability for data systems to support business processes in the pursuit of business goals degrades over time when data quality degrades. This results in more cost but less benefit. Attempting to fix these problems at the operational level will not work. Adding more data and data systems in an effort to "fix" these problems only makes them worse.

What will fix these problems is not more technology, more systems or more data, but Data Governance. Data Governance advances the twin goals of Data Quality and Data Reusability; data that is timelier, more accurate, more complete, more accessible, more useful and less costly.

Data Governance is not a technology function. It is driven by the business and forms a bridge between business management and technology providers. We need to evolve from Data Governance 1.0 to Data Governance 2.0. Data Governance 2.0 (DG2) is about the "WHO", the people within the organization necessary to achieve Data Governance goals and objectives.

Our Data Governance Framework identifies the desired future-state needed to address the problems that compromise our information asset today. It documents the Department's Information Architecture. Taking our cue from the Practice Guidelines, we will align information management concepts, practices, and context. (White, 2014) This framework will lead us to Data Governance 2.0.

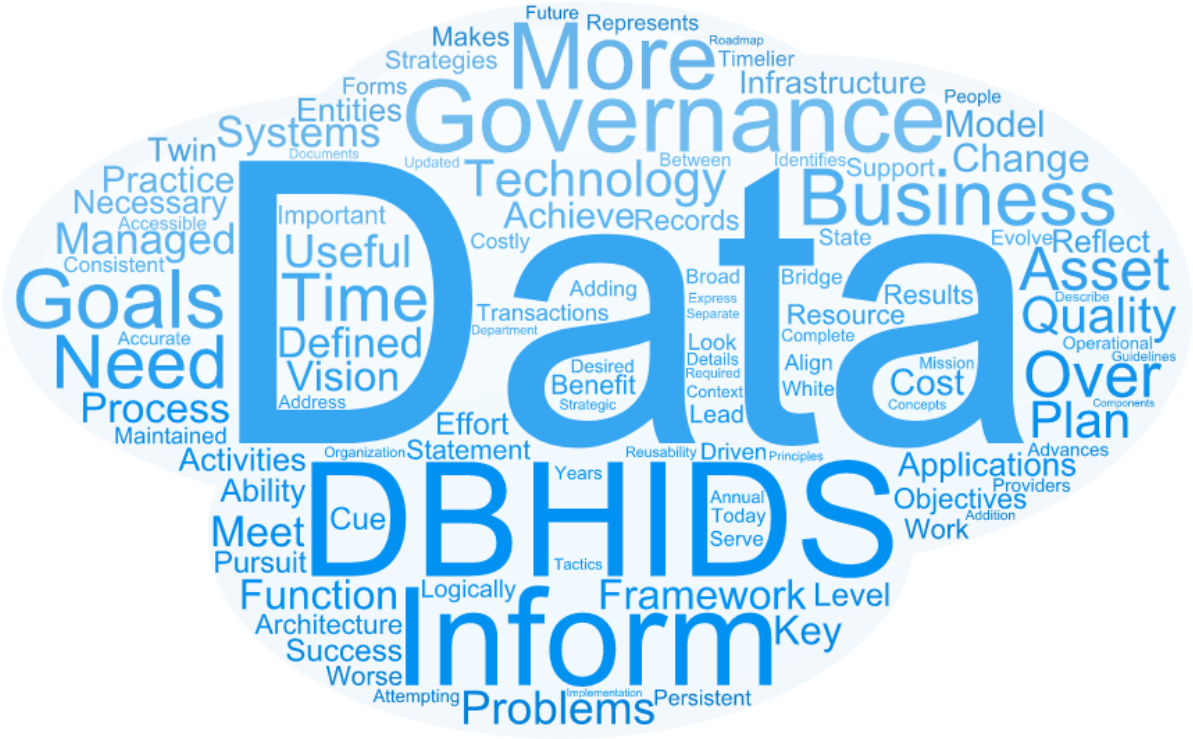
The **DBHIDS Data Governance Framework v2** includes the following strategic plan components, which will be updated every two to three years. This is the "what and why" that defines our activities.

- **The DBHIDS data governance mission statement**
- **The DBHIDS data governance vision statement**
- **Ten information architecture principles** that express the vision and inform the strategies
- **Three broad data management strategies** that serve as a roadmap to our goals
- **Ten data governance goals** that describe what success will look like

In addition, the framework includes a separate annual **implementation plan** that details the objectives and required tactics to achieve success. This is the how, who, when, and where of our activities.

PREAMBLE TO THE 2017 UPDATE

The Philadelphia Department of Behavioral Health and Intellectual disAbility Services (DBHIDS) Data Governance Framework Strategic Plan is over two years old and due for an update. In addition, it can be better organized. For this reason, version 2 represents a substantial rewrite.



The current DBHIDS Data Governance Framework consists of a Strategic Plan (conceptual), an Implementation Plan (logical design), and multiple project plans (physical instantiation) necessary for success. Version 1 of the Strategic Plan is twenty pages, supported by an eighteen page Implementation Plan with an additional forty-four pages of appendices. Moving forward, the strategic plan will be updated about every thirty months and the implementation plan will be updated annually.

The reasons for the update to the Strategic Plan include:

- We know more now than we did two years ago.
- Our organization mission is evolving.
- Our data management capabilities are growing.
- We have implemented things that did not exist before the original plan and based upon what we learned we can now evaluate their efficacy and adjust efforts moving forward.

Ongoing changes in technology is not listed as a reason for updating the plan. This is deliberate. Data governance is all about business use of data. The technology comes and goes but the need for understanding of the data is a constant. A strategic plan should describe the need to use appropriate technologies and stay current with developments but should not be prescriptive about specific technologies that can change over time.

Over the past year, analysts and business unit management were interviewed about the state of data management and information delivery in the department. Through these conversations, consistent complaints arise were identified concerning workload, including:

1. People identified as "data analysts" do not provide data fast enough to meet the needs and expectations of people identified as "research analysts".
2. Too much effort goes into cleaning and standardizing data for each request.
3. Information that management expects to be available requires too much time to produce.
4. Overall data quality is poor or inconsistent; too much effort goes in to verifying reliability.

The Enterprise Data Management unit investigated the complaints and determined that they in large part are the result of the following, respectively:

1. The department lacks data governance focus, and has neither fully identified the need nor assigned the responsibility for data quality.
2. The department has an imbalance in the types of analysts, with a significant number of self-identified research analysts, too few data analysts, and almost no business data analysts.
3. The department focuses on delivering final products without leveraging what was learned in creating the product.
4. The department does not take a data-driven, business intelligence approach to anticipating the organization's reporting needs.

A factor linking these causes to the complaints above is: ***placing responsibility for data management in research units.*** The organization made a faulty assumption that because research analysts were best prepared to evaluate the results of data collection and processing, that they would also be the best qualified to manage the data. Unfortunately, the data management goals of an organization are fundamentally at odds with the data needs of a researcher. We need to refresh the department's Data Governance Framework Strategic Plan so that it refocuses our strategy on meeting department data governance goals, not research goals. (Seiner R. T., 2005)

There is an interesting paradox to the research-centric approach. Anything that was remotely database related was assumed to be "technology" and there was no recognition that technologists are not data management professionals. Management assumed that if research analysts were not addressing data management concerns then network administrators and programmers must be doing so. To further exacerbate the situation, the research analysts became proxies for the necessary role of business data analyst, which further widened the chasm between business units and their data.

Data Management is Plumbing

"The society which scorns excellence in plumbing as a humble activity and tolerates shoddiness in philosophy because it is an exalted activity will have neither good plumbing nor good philosophy; neither its pipes nor its theories will hold water."

— John W. Gardner,
Excellence: can we be equal and excellent, too? (Gardner, 1961)

Using a Strengths, Weaknesses, Opportunities, and Challenges (SWOC) analysis, this Strategic Plan documents the desired data governance and data management goals of executive management, identifies why the current approach has been unsuccessful in meeting those goals, examines options for a restructured approach, and makes a recommendation for the approach most likely to succeed, all factors considered. It also identifies the human resource challenges to be faced by DBHIDS in trying to implement the recommended approach. Furthermore, this revised Strategic Plan is aligned with the DBHIDS Mission and Vision.

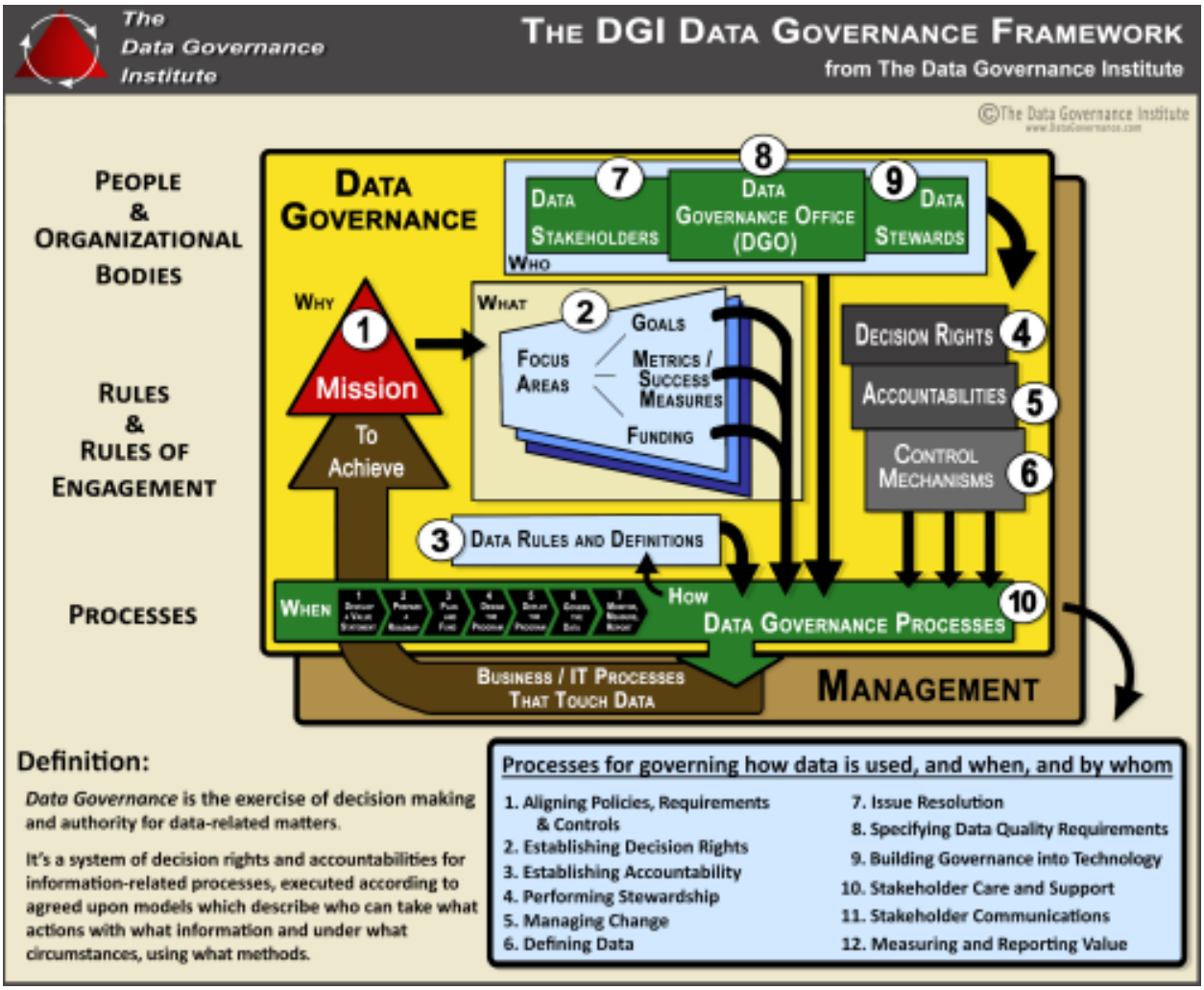
DBHIDS Mission Statement	**Data Governance Mission Statement**
The mission of the Department of Behavioral Health and Intellectual disAbility Services is to educate, strengthen, and serve individuals and communities so that all Philadelphians can thrive. (adopted January, 2017)	*Our data governance mission is to define and manage a quality data resource that enables DBHIDS to educate, strengthen, and serve individuals and communities to fulfill its mission.*
DBHIDS Vision Statement	**Data Governance Vision Statement**
We envision a Philadelphia where every individual can achieve health, well-being and self-determination. (adopted January, 2017)	*We envision an enterprise information architecture that provides secure, defined, quality data whenever and wherever needed in a cost-effective, reusable, and repeatable manner so that DBHIDS can realize its vision.*

THE NEED FOR A DATA GOVERNANCE FRAMEWORK

WHAT IS A DATA GOVERNANCE FRAMEWORK?

A Data Governance Framework is a logical structure for classifying, organizing, and communicating complex activities involved in making decisions about and taking action on enterprise data. (Thomas, ND) A data governance framework (DGF) documents the information architecture and data governance practices for an organization which guide its enterprise data management. It prescribes an approach to data governance, data management, and data architecture as well as the information technology to support the goals of data reusability and data quality. It identifies the roles and responsibilities throughout the organization.

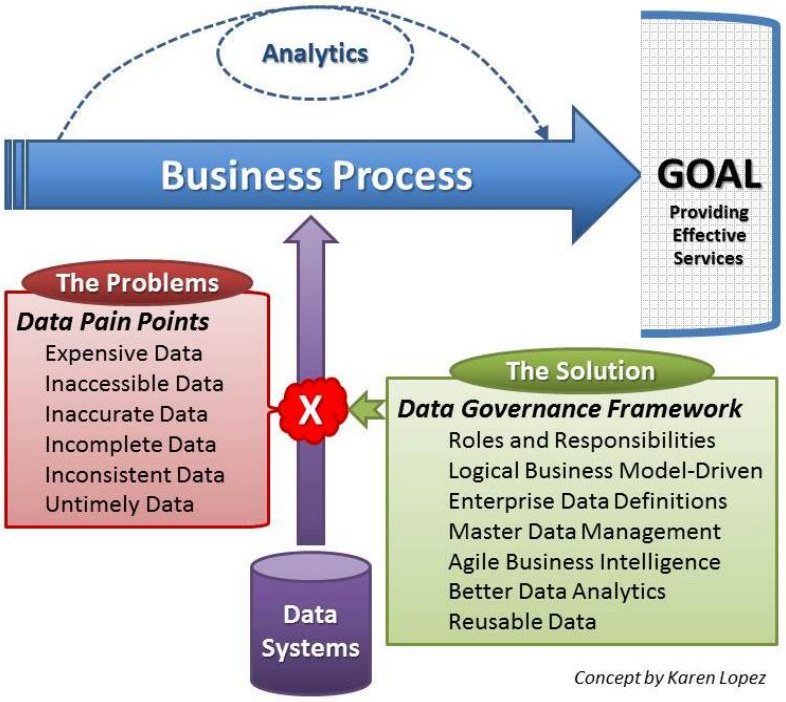
The Data Governance Institute has developed a model data governance framework that focuses on the roles and responsibilities of data governance, and these are the emphasis of the DBHIDS Data Governance Framework version 2. (Thomas, ND) We list these in Appendix E. However, our data governance framework cannot exist in a vacuum disconnected from our information and data management. For this reason, we extend our data governance framework to encompass information architecture and data management implementation activities.



DATA GOVERNANCE FRAMEWORK AS INFORMATION ARCHITECTURE

A data governance framework also embodies our organization's information architecture. It is how we maximize the value of our data by defining it, the use cases for it, and the roles and responsibilities for governing and managing it. The goal is to make sure that defined data of known data quality (fit for the intended purpose) is available to our staff whenever and wherever needed. The ultimate aim is to help our organization best meet its mission of improving population health for city residents.

While it is important to discuss the benefits to the organization that will be provided by an enterprise approach, it can be difficult to make these over-arching issues resonate with business management. While keeping the big picture of enterprise information management and data governance in perspective, we must demonstrate how the strategic plan addresses the pain points of the individual business units.



Practical Architecture

Moving forward, for each project, we must connect the enterprise activities required by this implementation plan to either the solution sought by the business or the avoidance of both seen and unforeseen problems; ideally, we will accomplish both.

John Ladley trademarked GAIP (Generally Accepted Information Principles™ – see **Appendix B**) to provide organizations with principles that they could use to guide data quality programs and initiatives. These address the value of data, the risk of not managing it properly, the quality of data and the need for accountability (in terms of roles and responsibilities) for both the data and its quality. (Ladley, 2012) The intersection of the concepts of data reusability and data quality is data governance.

Peter Drucker described how one of the benefits of investing in automated equipment is a reduction in non-producing time by improving quality ("that is, getting it right the first time"). He also predicted that large monolithic manufacturing plants and processes would not be able to compete with modular processes of known quality designed to work together to produce solutions (Drucker, 1992), analogous to the DBHIDS information architecture in support of data quality and data reusability.

Data is the fundamental building block of Digital Government. It is a critical resource and we must manage it as such. We must transform the practice of creating isolated islands of data to satisfy individual programs or units. We must manage a core of common data at the enterprise level. We must manage all data with common tools and methodologies. (Ladley, 2012) This will make it possible to use data management technologies to collect, publish, and maintain the integrity of critical data elements across multiple programs in a manner that is both efficient and responsive to business needs. Formal information architecture is essential to achieving this. (Seiner R. T., 2005)

Information architecture is a component or perspective of the enterprise architecture. Information architecture represents the reference architecture for an enterprise data management program. Reference architectures describe the vision, goals, objectives, principles, practices, standards, methodologies, and tools used in a particular technology domain within an organization. The data management domain encompasses the collection, definition, and maintenance of data and the development and presentation of actionable information derived from that data.

Enterprise Architecture

Enterprise architecture (EA) is a well-defined practice for conducting enterprise analysis, design, planning, and implementation, using a holistic approach at all times, for the successful development and execution of strategy.

EA applies architecture principles and practices to guide organizations through the business, information, process, and technology changes necessary to execute their strategies. These practices utilize the various aspects of an enterprise to identify, motivate, and achieve these changes.

Normally an organization's EA takes the form of a comprehensive set of cohesive models that describe the structure and functions of an enterprise. The individual models in an EA are arranged in a logical manner that provides an ever-increasing level of detail about the enterprise.

en.wikipedia.org/wiki/Enterprise_architecture

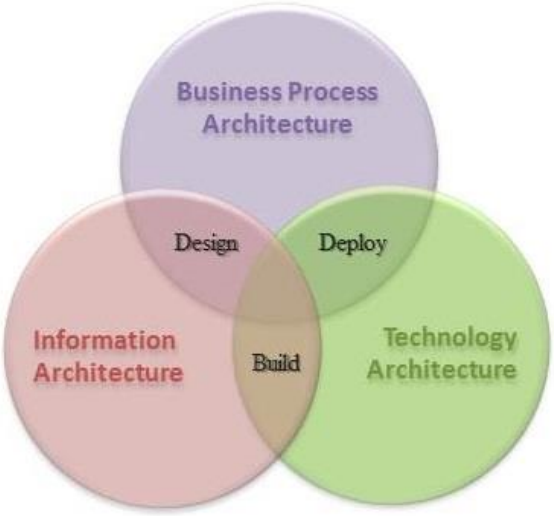
Relationship to Other Architectures

In order to place domain architectures such as the Data Governance Framework in context, it is essential to acknowledge one overarching relationship:

The DBHIDS Data Governance Framework is one of three domains (components) of an overall DBHIDS Enterprise Architecture.

This enterprise architecture consists of three related architecture domains:

- Business Process (the business operations)
- Information (the data), represented by this Data Governance Framework
- Technology (the hardware, network, software, and security platforms)



The Business Process Architecture provides the essential functionality of the business; what it means to be the business. The Information Architecture guides the development of data necessary for the Business Process Architecture; what it means to be of interest to

the business. The Technology Architecture provides the components required by the Information and Business Process Architectures.

Collectively, these domains define the solution architecture for a specific business problem. Of the three, Information Architecture is the least volatile and longest-lasting, as the record of items of interest to the business outlives changes to process or technology.

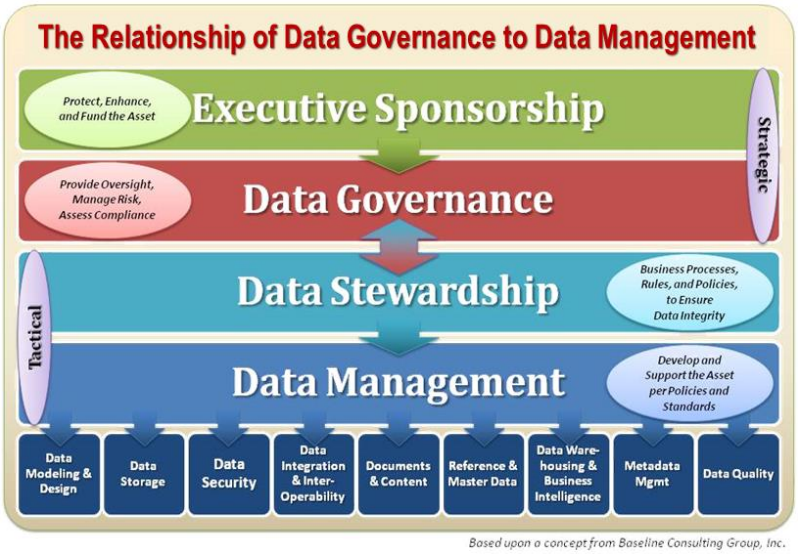
Without an understanding of these crucial relationships, business users and technologists cannot help but create the islands of disintegration that enterprise architecture is charged with preventing.

Data Governance and Its Relationship to Information Architecture

Information Architecture describes what, when, where and why; in other words, the domain of data management. Data Governance describes who and how; in other words, roles and responsibilities. The Data Governance Framework documents our information architecture within the data governance context of roles and responsibilities (see Appendix E).

Broadly speaking, Data Governance is the exercise of decision-making and authority for data-related matters. Formal Data Governance is a system that provides rules and policies proactively to enable ongoing efficient service delivery while addressing data quality issues as they are identified. Data Governance also refers to the organizational bodies, rules, decision rights, and accountabilities of people and information systems as they perform information-related processes. In other words, Data Governance is the identification of those with decision-making responsibility for data management. (Seiner R. T., 2005)

Data Governance is not a technology function. It is driven by the business and forms a bridge between business management and technology providers. ***Executive Sponsorship*** comes from the business. ***Data Governance*** and ***Data Stewardship*** represent collaborations of business subject matter experts and information architecture staff. ***Data Management*** is provided by various technologists. (Ladley, 2012)



The DBHIDS Data Governance Framework describes the DBHIDS Information Architecture as well as the roles and responsibilities necessary to implement it successfully. These concepts will be described more fully in the sections that follow. It is informed by a data strategy assessment produced by an assessment tool published by the Harvard Business Review that is included in **Appendix D**.

The Importance of an Architectural Framework

Architectural frameworks are a form of strategic planning. A good framework will document the current state of the organization as well as the desired state. It will plot the route for evolving from the current state to the desired state. In this, it provides its greatest value, as it becomes the way to "decide how to decide". It establishes the practices that will be followed to address categories of situations.

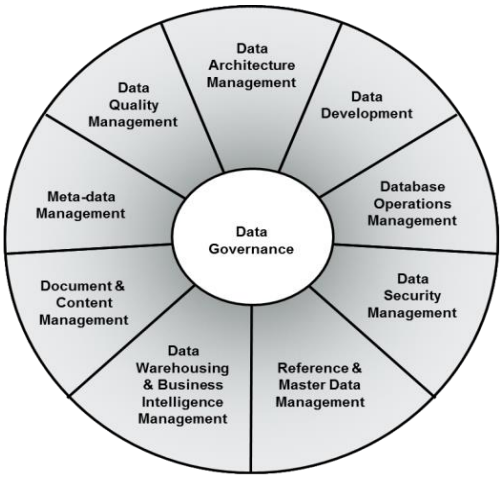
Organizations make tactical decisions every day. These can be as mundane as what to name something or as important as where to get authoritative information. Without a plan, the organization will not know when one of these decisions is misaligned with the desired state. A data governance framework provides a rational description of both our information architecture and how we will implement it. It serves as a guide for decision-making around data, data technologies and data management processes.

The Relationship between a Data Governance Framework and the DAMA DMBOK

DAMA (the Data Management Association) is a not-for-profit, vendor-independent, international association of technical and business professionals dedicated to advancing the concepts and practices of information resource management and data resource management. The "body of knowledge" about these domains is quite large and constantly growing. To respond to this challenge DAMA International provides the DAMA Guide to the Data Management Body of Knowledge, or DAMA DMBOK, as a definitive introduction to data management.

The DAMA DMBOK defines a standard industry view of data management functions, terminology and best practices, without detailing specific methods and techniques. While DAMA DMBOK is not a complete authority on any specific topic, it will point readers to widely recognized publications, articles and websites for further reading.

DAMA has identified ten distinct data management domains and represents them in the DMBOK Wheel, as shown here. Data Governance is one of the domains. It overarches the other nine domains, providing for coordination between them and facilitating communications and planning.



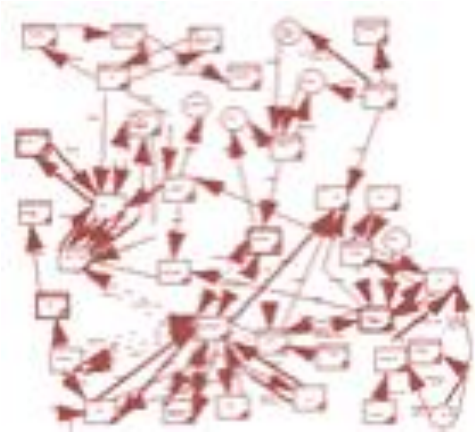
Copyright © by DAMA International

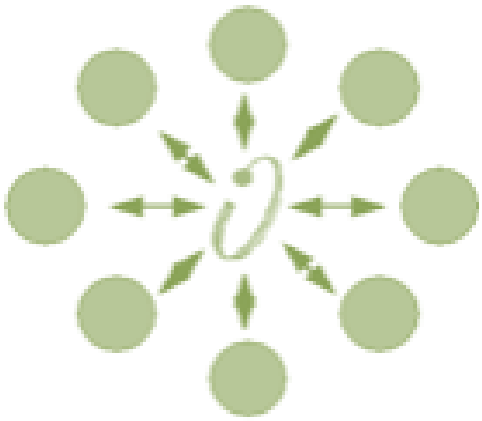
The DMBOK Wheel illustrates that each of the nine management domains has equal value; no one discipline is more important than any other. It also illustrates, however, how each of the nine management domains is influenced by the organization's data governance processes. It identifies data governance as the central process with which all other data management activities must interact. (DAMA International, Inc., 2014)

This DMBOK Wheel is not itself a data governance framework. It serves as an organizing scheme for discussing the interrelated data management disciplines and their dependence upon effective data governance. An organization still needs a data governance framework that reflects its principles and goals while addressing these knowledge domains.

Data Quality Represents Our Mission; Data Reusability Represents Our Vision

Data quality is defined as "Data Fit for Purpose". This includes an enterprise-approved definition, an identified authoritative source, and an understanding of the general accuracy and quality of the data. (Wilder-James, 2013) Data reusability is the refocusing of data management from a "sharing data by moving it around" approach to a "sharing information through reusable data" approach.





Sharing data by moving it around ***Sharing information through reusable data***

The purpose of the DBHIDS Data Governance Framework is to improve data quality and drive data reusability to meet the Department's strategic and operational needs.

Data quality is not one-dimensional. Accuracy is only one aspect. Wang and Strong suggest that there are four broad categories of data quality: Intrinsic, Contextual, Representational, and Accessibility. These categories have fifteen more-detailed dimensions. (Wang, 1996)

Malcolm Chisholm, the "Father of Reference Data Management", further expresses that data has context and is fit for some purposes but not for others.

> "We must not blame quality on the improper choice of data for a process. For example, if an analyst has metadata available to him and he doesn't do a good job with due diligence — i.e., he makes assumptions about the data that produce bad results in his analytics — then it's not the data's fault.
>
> "…For data to represent what it purports to represent, you must know what it's supposed to be representing. That's metadata — and what's purported has to match the actual data values. Of course, the extent to which data satisfies a specific requirement is context-dependent." (Chisholm, 2017)

Data reusability represents an evolution beyond the basic concept of data sharing. Data reusability supports sharing of high quality information as everyone consumes the same version. Data sharing contributes to poor data quality as multiple versions circulate throughout the organization. An effective information architecture approach enhances data reusability and data quality by eliminating stand-alone data sharing activities.

In its report, "***NATIONAL INFORMATION ARCHITECTURE: Toward National Sharing of Governmental Information***", the National Association of State Chief Information Officers (NASCIO, nee NASIRE) summarizes that "The business case for such an architecture rests on four foundations." These four foundations are the ability to positively impact: Accuracy, Completeness, Cost/Expense, and Timeliness. (NASIRE, 1999) We have identified two additional foundations: Accessibility and Usefulness.

Collectively, these represent six drivers for our data governance goals and objectives.

Reusable Data is More Accurate

Reusable data increases the reliability of data transactions. Standardized lookup tables provide developers with a low-cost and consistent source of reference data to validate data entry according to department standards. Reusable master entity information, retrieved on demand, reduces the potential for user input errors and update anomalies that develop between redundant data sets. Multiple versions of the same data lead to inconsistencies that are difficult to reconcile. When the same data is accessed by multiple users, errors are found more quickly.

Reusable Data is Timelier

Applications can make updates to Master Data available to all stakeholders immediately. There is no need to update disparate systems manually, thus eliminating workflow bottlenecks. As data is integrated for reuse, reporting and analysis can take the form of self-service. Turnaround time for new reports and requests for information is greatly reduced. Latency between data collection and the ability to report on it is also greatly reduced.

Reusable Data is More Complete

Reusable data enables stakeholders to access aggregate master data from a single access point. They will not need to work with multiple systems just because multiple business units manage those records. Developers can write applications to recognize, in an intelligent way, dependent processes across unit and line-of-business borders. Logical workflow can be incorporated into these applications to capture and maintain all related information.

Reusable Data is Less Expensive

There are hundreds of data tables in DBHIDS databases that duplicate data available elsewhere. These tables contain information as basic as county codes for lookup validation, or as critical as demographic data for entitlement programs. Centralized management of universal information reduces the costs of creating, maintaining and reconciling multiple containers of the same information.

Reusable Data is More Accessible

An Enterprise Business Architecture Model and corresponding metadata provide stakeholders with the roadmap and the mechanism to interoperate electronically. Constraints on data sharing, whether valid or merely perceived, are resolved as a part of the Business Architecture Model creation and maturation process. Applications can access Reusable data to the extent permitted by established business rules and legal requirements.

Reusable Data is More Useful

Decision making within the organization is improved through the use of self-service reporting and key performance indicator (KPI)-based dashboards. This information can come from multiple sources across the enterprise. To be successful, these capabilities should be built upon a stable enterprise data warehousing environment that ensures that consistent answers are retrieved regardless of report mechanism or timing of the request. (Inmon, Strauss, & Neushloss, 2008)

Collectively, these represent the value of the Data Governance Framework and comprise the strategic drivers for the DBHIDS Information Architecture in support of Data Reusability.

The DBHIDS Data Governance Framework (DGF) represents the information architecture for the Department and guides its enterprise data management. It prescribes an approach to data governance, data management, data architecture, and information technology to support the goal of data reusability.

The DGF "pre-decides" many of the routine (obvious) decisions by prescribing best practices and standards. It describes a governance structure for identifying novel, emergent and good practices that eventually evolve into best practices. The framework identifies the goals, objectives and principles that guide decision-making about these practices.

The DGF represents a commitment to information architecture as a long-term strategic initiative to enable data quality and data reusability. This architecture forms the foundation for collecting, storing, managing, controlling privacy of, and providing access to enterprise data to meet business needs.

By following it, stakeholders will have access to more useful information, as they:

- Collect data once but use it often,
  improving data accuracy
- Store data more effectively
  for a timelier and more complete information picture
- Reduce or eliminate costs associated with
  data collection, storage and error correction
- Improve access to information while
  better protecting the privacy of individuals

DATA GOVERNANCE 1.0 VERSUS DATA GOVERNANCE 2.0

At the same time that DBHIDS is updating its Data Governance Framework from version 1 to version 2, the data profession has evolved its definition of data governance. The original DBHIDS Data Governance Framework was based upon what is now considered Data Governance 1.0 (DG1). DG1 was process-oriented. It was about control and authority. It was about the "How", the responsibilities necessary to achieve Data Governance goals and objectives. The following definitions reflect DG1.

> **Data governance** (DG) refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. (TechTarget)

> **Data governance** is the practice of organizing and implementing policies, procedures and standards for the effective use of an organization's structured/unstructured information assets. (Karl, 2008)

> **Data Governance:** The execution and enforcement of authority over the management of data assets and the performance of data functions. (Seiner R. , 2006)

Version 2 of the DBHIDS Data Governance Framework has embraced the precepts of Data Governance 2.0 (DG2). Compared to DG1, DG2 is about the "WHO", the people within the organization necessary to achieve Data Governance goals and objectives. Consider the following definition:

> **Data governance** is the decision-making process that prioritizes investments, allocates resources, and measures results to ensure that data is managed and deployed to support business needs. (Dychê, 2008)

The Data Governance Institute has evolved its definition to encompass this people-centric approach:

> "Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods." (Thomas, ND)

This last is a general, all-purpose definition of Data Governance, focused at the mid-level managers who must come together to make cross-functional decisions, set policies, and execute them. The Data Governance Institute believes that it properly focuses on rules of engagement components of a data governance framework, thereby accommodating if not enabling a more participative, consensus-based approach than the authoritarian process-centric view of DG1. The ultimate purposes of a successful data governance program are to improve enterprise data quality and data reusability.

Our department recognizes the need for formal data governance to address data quality and to make data available fit for purpose. One motivation for the creation of the original Data Governance Framework Strategic Plan was to address data governance roles and responsibilities. (DBHIDS, City of Philadelphia, 2016) The refresh of the Data Governance Framework Strategic Plan is intended to apply what has been learned from initial efforts and develop new objectives and strategies where appropriate.

TRANSITION TO DATA GOVERNANCE FRAMEWORK VERSION 2

To best understand the evolution of the Data Governance Framework v2, it is helpful to compare the data management landscape of the department prior to 2015 with its state in 2017.

THE DATA MANAGEMENT LANDSCAPE PRIOR TO 2015

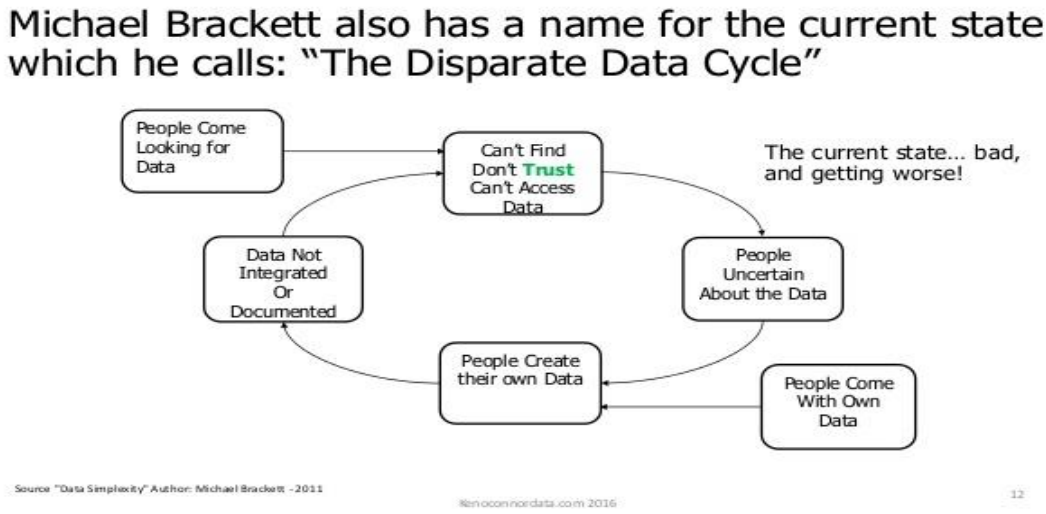
DBHIDS is a diverse organization. Prior to the adoption of the Data Governance Framework Strategic Plan in 2015 it had a decentralized approach to data management, even for data that is of use across organizational boundaries. Most data systems were developed "bottom-up" to meet business unit needs. Information management resources were deployed to meet the needs of those business units, and were funded by those units. Funding was often determined by various federal or state programs which placed either real or perceived restrictions on data management and use. Even when data was common between business units, it was addressed by units exchanging data from time to time as needed without attempting to agree on definition, structure or common management.

The department addressed data quality problems by creating several large centralized groups of research analysts that were charged with knitting together data and cleansing it every time there was a request for a more integrated view. These research analysts became subject matter experts in the data of various business units yet did not have any day-to-day organizational responsibility to those units. Worse, they did not provide guidance to the multiple independent data collection development teams to help standardize the data. We lacked a "Network-centric Paradigm". (Crawford, Hasan, Warne, & Linger, 2009)

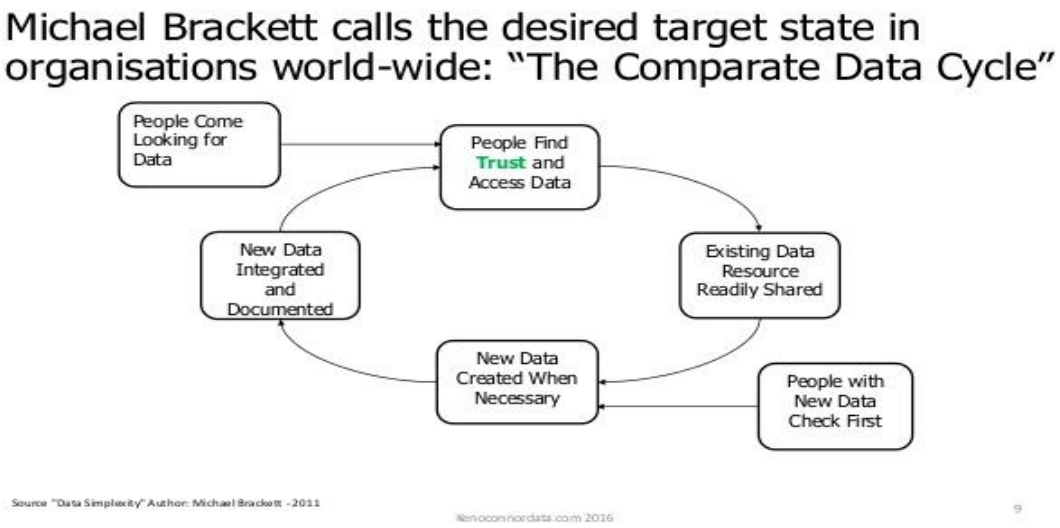
DBHIDS Developers and analysts were hard-working, earnest and intelligent but have lacked data integration tools, methodologies and training. They were able to deliver meaningful value but were unable to keep up with an ever-increasing workload. In addition, as most analytical efforts were performed one-off without leveraging previous efforts, data quality problems were introduced into reports due to lack of consistent business rules and definitions. The data as a result of many of these efforts was maintained in collections of spreadsheets, often without documentation and unavailable for reuse.

Because of this, data efforts were seen as competitions in which units and analysts strive to be the one with the "correct" answer rather than assisting the organization to achieve better data quality for everyone. The department lacked a "comparate data cycle", a self-perpetuating cycle that reinforces the use of trusted data. (Brackett, 2011) The result of this is that a significant amount of energy goes into inadvertently making data quality worse. (Loshin, 2011) These factors encouraged the development of data siloes and discouraged cooperation between units; what Brackett refers to as a "disparate data cycle". The 2015 Data Governance Framework Strategic Plan inaugurated the transformation of these practices.

Michael Bracket's Current State Model: **The Disparate Data Cycle**



Michael Bracket's Desired State Model: **The Comparate Data Cycle**



STATUS OF DBHIDS DATA GOVERNANCE STRATEGIES, VERSION 1

The following nine data governance strategic goals and ten data governance strategic objectives were identified in Version 1 of the Implementation Plan. In the chart that follows, Blue indicates that a strategic objective has been substantially realized; Green indicates it is well underway; Yellow indicates some progress; and, Red indicates no notable progress. The only objective without notable progress is "Provision data for those that need it as required."

2015 Strategic Goals

1. Create an information-centric and informed organizational culture.
2. Establish a data governance program to provide accountability for information assets.
3. Provide for effective and appropriate information security.
4. Improve the quality and usefulness of information by making it timelier, more accurate, more complete and more accessible.
5. Reduce the costs of managing information.
6. Share data through reusable processes; reuse data through shared processes.
7. Provide self-service business intelligence capabilities.
8. Develop enterprise-class data management staff.
9. Adopt enterprise-class data management tools.

2015 Strategic Objectives

1. Establish within the Office of the CIO a DBHIDS-wide data management office to address data governance, data architecture, data integration, and business intelligence and analytics.
2. Implement the Data Governance Framework
3. Achieve compliance with information architecture policies and standards
4. Develop new draft policies and standards for consideration by the DSC
5. Manage the enterprise business architecture model and business glossary
6. Provide enterprise-level metadata management for business and technical users
7. Manage master data management domains
8. Manage the data integration and persistence layer for the department
9. Provision data for those that need it in the manner required
10. Evangelize the Data Governance Framework principles to the department

Relationship of Data Governance Strategic Objectives to Goals

STRATEGIC OBJECTIVES STRATEGIC GOALS		1	2	3	4	5	6	7	8	9
1.	***Establish the Office Enterprise of Data Mgmt.***	√	√	√	√	√	√	√	√	√
2.	***Implement the Data Governance Framework***	√	√		√	√				
3.	***Achieve compliance with policies and standards***	√	√	√	√	√				
4.	***Develop new draft policies and standards***	√	√	√	√	√				
5.	***Manage enterprise LDM and business glossary***	√	√		√	√	√	√		
6.	***Provide enterprise metadata management***	√	√		√	√		√		
7.	***Manage master data management domains***	√	√		√	√	√			
8.	***Manage data integration & persistence layer***		√	√	√	√	√		√	√
9.	***Provision data for those that need as required***	√		√	√		√	√	√	
10.	***Evangelize the Data Governance Framework***	√	√						√	

THE DATA MANAGEMENT LANDSCAPE IN 2017

As a result of the efforts to implement the Data Governance Framework 1.0, the data management landscape in 2017 is radically different. DBHIDS is still a diverse organization. It has a centralized enterprise data management focus while incorporating a federated model for data stewardship and business analysis. Most new data system development is reviewed first to leverage existing capabilities and align with department data requirements. Business units recognize the difference between data that needs to be collected to document transactions and reports and analysis that are derived from that data. Most new requests for data are reviewed to leverage existing data, avoiding redundant effort, as well as to identify new opportunities to build sustainable solutions. While master data management has not yet been implemented, data stewards have been identified for various MDM domains and have begun identifying and reconciling business rules required for each domain.

A Strengths, Weaknesses, Opportunities, and Challenges analysis was completed and is included as **Appendix C**.

Strengths include the establishment of the Office of Enterprise Data Management, the establishment of the Data Governance Executive Board and Data Stewardship Council, the creation of master data management domain data stewardship committees, the initiation of the enterprise data warehousing, master data management, and business intelligence tool initiatives, and the identification of guidelines for the coordination of data analyst activities. The most important strength continues to be a commitment by executive management to support enterprise data governance, data quality, and data management programs.

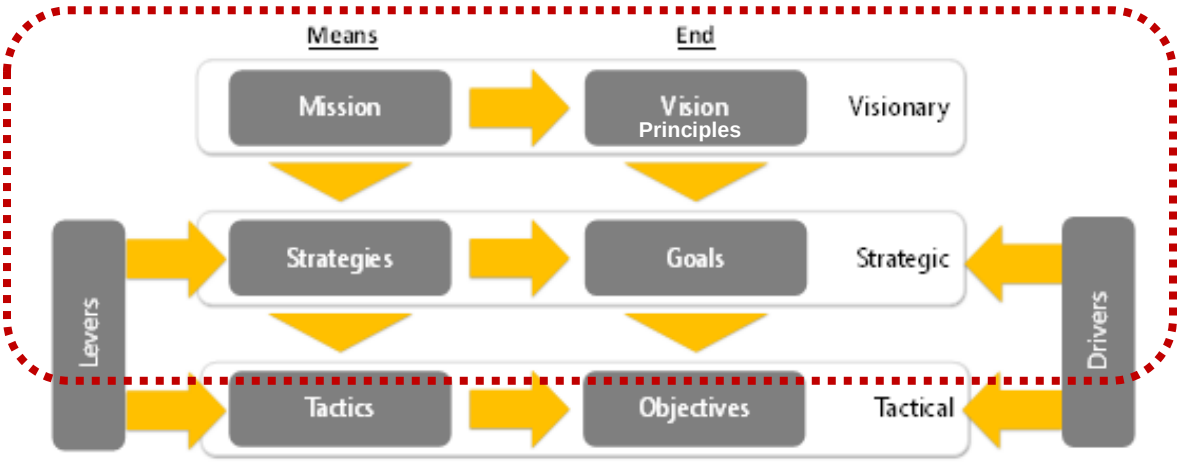
Unfortunately, there are several significant **weaknesses** that must be addressed. There is a continued treatment of some data requests as custom research projects, resulting in redundant effort, no sustainability, and poor data quality. Some business unit managers do not accept their role in prioritizing data requests and projects. We still rely too much on outside vendors and contractors for data architecture and design decisions that should be owned by the organization.

There are numerous **opportunities** present in 2017. The loss of key staff has highlighted the danger of relying on "hero-based" analysts and has steered the organization towards a formal, enterprise approach. We have demonstrated value of the information architecture to executive management through pilots and small-scale projects even as large-scale initiatives were getting underway. We have filled many staffing needs but there remain opportunities to add needed skills and capacity. We need to refine the data request process and make sure that it is in use throughout the organization.

Many **challenges** also remain. Our analyst organizations are still too "research analyst" heavy, which does not provide the right skills to improve data quality. Our new enterprise data management efforts are often slowed or delayed due to the day-to-day needs of the organization for data produced through the less-efficient legacy processes. The capabilities and resources of the City of Philadelphia's central information technology office (OIT) do not always align with our needs.

STRATEGIC PLAN COMPONENTS

Our Data Governance Framework organization is based upon a business motivation model used in enterprise architecture efforts. The chart below is from Craig Martin of Design Thinking in Business. It identifies both strategic and tactical (implementation) components. To it, we have added the concept of Principles that further define the Vision. The components relevant to this strategic plan are encircled and are defined below. Tactics and Objectives are defined in the DBHIDS Data Governance Framework Implementation Plan.



Business Motivation Model (Martin, 2013)

These are the definitions of the components. The DGF Strategic Plan descriptions are in the next section.

DATA GOVERNANCE MISSION STATEMENT

A mission statement is the definition of an organization's purpose. It documents the reasons why stakeholders should support the organization. (Bryson, 2011) To assure its relevance, the DBHIDS data governance mission is aligned with the DBHIDS agency mission:

The mission of the Department of Behavioral Health and Intellectual disAbility Services is to educate, strengthen, and serve individuals and communities so that all Philadelphians can thrive.

(DBHIDS Mission Statement, adopted January, 2017)

DATA GOVERNANCE VISION STATEMENT

A vision statement is the description of the code of behavior to which an organization aspires. (Bryson, 2011) To assure its relevance, the DBHIDS data governance vision is aligned with the DBHIDS agency vision:

We envision a Philadelphia where every individual can achieve health, well-being and self-determination.

(DBHIDS Vision Statement, adopted January, 2017)

Our data governance vision is further embodied by a set of data architectural principles to which we adhere and which informs our strategy.

Data Architecture Principles

An architecture principle is a comprehensive and fundamental law, doctrine, or assumption that provides overarching guidance for development of a solution. A good architecture principle is not outdated by advancing technology and, more importantly, provides objective reasons for advancing it instead of alternatives. The ten information architecture principles represent the vision that guides the identification of goals and objectives for our information architecture and the formation of strategies to achieve them. (Ladley, 2012)

Data Governance Goals

Goals describe concrete, action-oriented targets that categorize and focus information management efforts. The ten data governance goals are the heart of the department's information architecture. Each goal has one or more objectives that align to the information architecture principles.

Data Management Strategies

Strategies are the means by which an organization intends to accomplish a goal or objective. A strategy summarizes a pattern across policies, programs, projects, decisions, and resource allocations. (Bryson, 2011) The three broad data management implementation strategies will guide the implementation plan tactics to meet our objectives and achieve our goals.

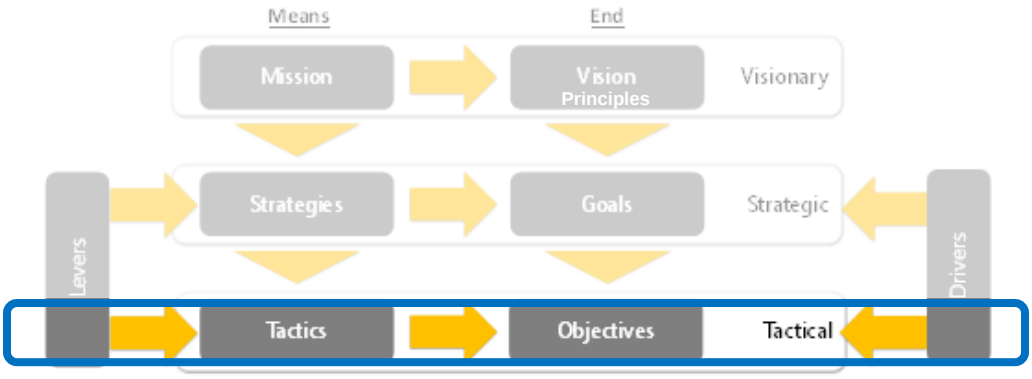
Information Architecture Drivers

A business driver is a resource, process or condition that is vital for the continued success and growth of a business. The six critical information architecture drivers are the impetus for our data governance goals and objectives.

Data Architecture Levers

A business motivation lever is a capability or design pattern that can influence policies, programs, projects, decisions, and resource allocations to be more effective and useful to the organization. Our data architecture levers facilitate and make more effective our data management strategies. These levers are embodied in our information service delivery use cases and our conceptual information architecture schema.

Information architecture drivers were described in a preceding section. The specifics of the remaining strategic components are described in more detail in the sections that follow. Tasks and objectives in the tactical layer are described in detail in the Data Governance Framework Implementation Plan.



THE DBHIDS DATA GOVERNANCE FRAMEWORK STRATEGIC PLAN

DBHIDS DATA GOVERNANCE MISSION

This is the DBHIDS data governance mission statement:

> *Our data governance mission is to define and manage a quality data resource that enables DBHIDS to educate, strengthen, and serve individuals and communities to fulfill its mission.*

DBHIDS DATA GOVERNANCE VISION

This is the DBHIDS data governance vision statement:

> *We envision an enterprise information architecture that provides secure, defined, quality data whenever and wherever needed in a cost-effective, reusable, and repeatable manner so that DBHIDS can realize its vision.*

DATA STRATEGY FORMULATION

In order to fulfill our mission and realize our vision, we must be able to deliver what Leandor DalleMule and Thomas H. Davenport refer to as a Single Source of Truth (SSOT). (DalleMule & Davenport, 2017) The SSOT represents the secure, defined, quality data required to support the DBHIDS vision. This requires centralized data governance and data management; what DalleMule and Davenport designate as a defensive data strategy. Not having an SSOT can lead to chaos:

> *"The SSOT is a logical, often virtual and cloud-based repository that contains one authoritative copy of all crucial data, such as customer, supplier, and product details. It must have robust data provenance and governance controls to ensure that the data can be relied on in defensive and offensive activities, and it must use a common language—not one that is specific to a particular business unit or function. Thus, for example, revenue is reported, customers are defined, and products are classified in a single, unchanging, agreed-upon way within the SSOT."*

However, to fully support the mission of DBHIDS, we must also be agile and provide data to various business units for distinct audiences presented in different contexts. DalleMule and Davenport describe this as Multiple Versions of Truth (MVOT). This is the quality data resource that can be used throughout DBHIDS to fulfill its mission. This requires agile, decentralized analytic and visualization capabilities supported by defined data; what DalleMule and Davenport designate as an offensive data strategy. Multiple versions of the truth, derived from a common SSOT, support superior decision making.

> *"An SSOT is the source from which multiple versions of the truth are developed. MVOTs result from the business-specific transformation of data into information—data imbued with 'relevance and purpose.' Thus, as various groups within units or functions transform, label, and report data, they create distinct, controlled versions of the truth that, when queried, yield consistent, customized responses according to the groups' predetermined requirements."*

Defensive and offensive strategies have different attributes.

The Elements of Data Strategy

	DEFENSE	OFFENSE
KEY OBJECTIVES	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve competitive position and profitability
CORE ACTIVITIES	Optimize data extraction, standardization, storage, and access	Optimize data analytics, modeling, visualization, transformation, and enrichment
DATA-MANAGEMENT ORIENTATION	Control	Flexibility
ENABLING ARCHITECTURE	SSOT (Single source of truth)	MVOTs (Multiple versions of the truth)

FROM "WHAT'S YOUR DATA STRATEGY?" BY LEANDRO DALLEMULE AND THOMAS H. DAVENPORT, MAY–JUNE 2017 © HBR.ORG

DalleMule and Davenport provide an assessment tool to determine whether an organization's strategy is more defensive or more offensive. (**hbr.org/2017/05/whats-your-data-strategy**) Our overarching data strategy is considered "defensive". We emphasize regulatory compliance and data control, security, privacy, integrity, and quality. This strategy is supported by a Single Source of Truth (SSOT). At the same time, we need flexibility meet information needs of the business from the SSOT; this requires the ability to support Multiple Versions of Truth (MVOT). See **Appendix D** for more information.

The SSOT-MVOT model provides the greatest value to an organization. It addresses the concerns embodied by a defensive strategy yet provides the benefits of an offensive strategy. It is also the most difficult for many to conceptualize, let alone operationalize. There are not many organizations that are able to tightly control data yet enable to to be used flexibly. Again from DalleMule and Davenport:

> *"Although the SSOT-MVOT model is conceptually straightforward, it requires robust data controls, standards, governance, and technology. Ideally, senior executives will actively participate on data governance boards and committees. But data governance isn't particularly fun. Typically, enterprise CDOs and CTOs lead data and technology governance processes, and business and technology managers in functions and units are the primary participants. What's critical is that single sources of the truth remain unique and valid, and that multiple versions of the truth diverge from the original source only in carefully controlled ways."*

This has implications for our implementation approach, addressed in the Data Governance Framework Implementation Plan. It informs our Information Architecture Principles which further expand our vision. It influences our data governance goals. It frames our data management strategies. These are described in more detail in the sections that follow.

DBHIDS INFORMATION ARCHITECTURE PRINCIPLES

1. **Information architecture is the reflection of the business; it is not just a technology domain.**
   This principle is critical to both a successful data governance effort and to individual data management projects. When the business abdicates responsibility for information architecture and data governance to information technologists, it leads to the creation of data silos, disparate data, poor data quality and a focus on activity over value. The business must partner with technologists in data governance and information architecture efforts. (Seiner R. T., 2005)

2. **The identification and definition of data attributes must involve the business.**
   When the business does not lead this effort, there is a loss of understanding over time that can neither be fixed nor replaced through the efforts of information technologists alone. For business-critical data elements, the respective business units must identify data ambassadors within the organization that can maintain the integrity of data definitions and approve the appropriate use of data for the desired purpose. (Seiner R. , 2006)

3. **Data is an organizational asset and must be managed with an enterprise perspective.**
   Once the business has taken responsibility for its role in data governance and data ambassadors are identifying and defining data attributes, the data must be managed at an enterprise (central) level. Data management decisions cannot be made at the system or program level. Because the data is an enterprise asset, decisions regarding how it is managed must also be made at the enterprise level. (Council for Information Advantage, 2010)

4. **Data that is common to more than one business unit must be defined through consensus by representatives of those business units.**
   It is essential that data that is used by more than one business be defined by representatives of all of the business units. When units are not represented in decision making, their specific needs may not be reflected. This is what leads to units creating their “own” versions of common data, as they are unable to use the “official” data. This process of business participation in the definition of common data is called data governance. (Wilder-James, 2013)

5. **The value of data to the enterprise is in its fitness for reusability, not its exclusivity.**
   To process data and exploit only the result of the calculation is short-sighted. Even worse is to lock it away. The practices and tools of effective data management cannot stand alone in the data ecosystem. They must rely on and support the reusability of data. The organization benefits when both data management efforts and results form a platform for future discovery and innovation. As big data, analytics and Web 2.0 grow in maturity and adoption, there will be a rising need to support exchange, collaboration and reuse around enterprise data. (Loshin, 2011)

6. **The value of data management staff is in its ability to build high quality, reusable data assets.**
   Data gatekeepers may perceive that they provide value to the organization by hoarding the data. This is exacerbated when institutional knowledge is often locked away in the memory of these individuals. When the individual leaves, the organization suffers the consequences. A mature data management staff adopts as its mission the commitment to make it easier for the business to gain access quickly to documented and defined information of known quality. (Ladley, 2012)

7. **Different information use cases require different data management solutions.**
   The technology necessary for processing transactional data is significantly different than the technology necessary for analytical processing or producing KPIs. The format of the data in these environments will be different. The security concerns for the data will be different. Each of these environments in turn is significantly different than one that is responsible for managing master or reference data or one responsible for storing data historically. (Inmon, Strauss, & Neushloss, 2008)

8. **In order to be sustainable, physical data stores must be governed by a conceptual understanding of the enterprise, captured in a business architecture model.**
   A Business Architecture Model (BAM) is not a database design. It represents the authoritative definition of data entities (people, places, things, events, etc.) and their attributes (characteristics) along with the relationships between the data entities (e.g. A Provider provides one or more Services, but must provide at least one). The BAM captures the business rules that govern data. This model is used to produce both logical and physical data models for specific solutions. (Hay, 2011)

9. **The purpose of a data management organization is to produce a data product that meets the information needs of the business commensurate with the investment made by the business.**
   In the same way that the business has an obligation to help define data and corresponding business rules, the data management organization has an obligation to implement solutions consistent with those definitions and rules using sound technology practices. Technologists must bring issues to the attention of the business that can have an adverse effect on data quality; the business will decide the priority and the appropriate investment for resolving those issues. (Brackett, 2011)

10. **Information architecture must be formal and proactive; it cannot be improvised or reactive.**
    Data must be defined, both technically and from a business perspective. Business rules must be defined. Data processes must be documented. Data quality issues must be documented. This documentation must be maintained in an organized manner and be accessible for those that require it. The enterprise must be able to measure both the quality and reusability of its data. When information is urgently needed, it is not the time to create an architecture approach or determine a data management strategy. (Loshin, 2011)

DBHIDS DATA GOVERNANCE GOALS

1. **Create an information-centric and informed organizational culture.**
   Becoming an information-centric organization requires substantial cultural change. Business and technology staff must become aware of the need, educated in the process and then empowered to approach information management with an enterprise viewpoint. Information architecture staff must evangelize and educate employees in both the value of this approach as well as how to implement it.

2. **Continue and strengthen the data governance program by adopting a centrally-managed data and information request process.**
   Data Governance is an approach to providing rules and policies proactively to define and manage data. Business data ambassadors and data architecture staff work together under management oversight. A centrally-managed data and information request process must engage the Data Stewardship Council and the Data Governance Executive Board so that they better understand the nature of information requests.

3. **Provide for effective and appropriate information security.**
   Information security is a multi-dimensional domain. It encompasses the Confidentiality of the data (protection), the Integrity of the data (accuracy and non-repudiation) and the Availability of the data (functionality). It is addressed through policies and procedures, education and awareness, encryption and access controls, and, vulnerability monitoring and auditing. It requires cooperation by business users, technologists and information security professionals.

4. **Improve the quality and usefulness of information by making it timelier, more accurate, more complete and more accessible.**
   Data management must be agile to meet business needs without making the data fragile and therefore unfit for use. This requires advance planning to leverage efforts to locate, define and integrate data one time but benefit from those efforts many times. By creating a catalog of reusable data – master, reference, operational and historical – all future efforts benefit. Once properly constructed, this complete and accurate data is available to more users and available to them more quickly.

5. **Reduce the costs of managing information.**
   There are obvious cost efficiencies achieved through better data management by the elimination of duplicate technology purchases, nor recreating data that already exists and not reinventing processes. There are even greater savings realized by eliminating the out-year and downstream maintenance of these inefficient processes and the decoupling of data use cases. Another substantial yet difficult to quantify cost savings is the elimination of data quality problems that lead to poor decision making.

6. **Reuse data through shared processes; share data through reusable processes.**
   Reusable data is data that has been integrated and published from a central store or repository, such as a data warehousing environment or a master data management platform. This data is best accessed through shared (common) processes implemented in an enterprise data integration environment. Other data requires access in real time so that it can be shared between transactional systems as needed. This is done through web services (reusable processes) implemented by each system.

7. **Provide self-service business intelligence capabilities.**
   The twentieth century model of business intelligence (technology worker-based) relied on a large IT staff creating reports for a small group of report consumers. The twenty-first century model for business intelligence (knowledge worker-based) provides for self-service, ubiquitous reporting capabilities, and access to quality data by and for all staff. The IT staff is responsible for data integration, providing access to data sources, and supporting a self-service business intelligence platform. Business data ambassadors are responsible for approving the business definitions of the data and identifying requirements and metrics. End users create reports as needed.

8. **Develop enterprise-class data management staff.**
   The skills required for data management in the twenty-first century are significantly different than those that were required in the twentieth century. Mainframe environments lent themselves to assembly-line skill delineation and data was kept in silos by design. Today, data management professionals need to be generalists and have the ability function as business analysts, data architects, data integration developers, business intelligence developers and database administrators as needed.

9. **Adopt enterprise-class data management tools.**
   Too often, data management tools are selected randomly, due to personal preference or perceived cost benefits, without considering the needs of the organization or the impact of using the wrong technology. The objective is not to select the “best tool” or the “least expensive tool” but to select a suite of tools that meet all of the needs of the entire organization, work well together and can be implemented and used for a reasonable investment of both money and staff time.

10. **Redefine the roles of data-related analysts in the organization to better match analyst capabilities to organization needs.**
    We must separate research analysts from data management responsibilities and identify the business data analysts and data analysts required by the organization. The organization is still too “research analyst” top-heavy and must identify or acquire business data analysts and BI specialists that can better meet business unit demand for data. The organization too easily slips back into “old data habits” under stress.

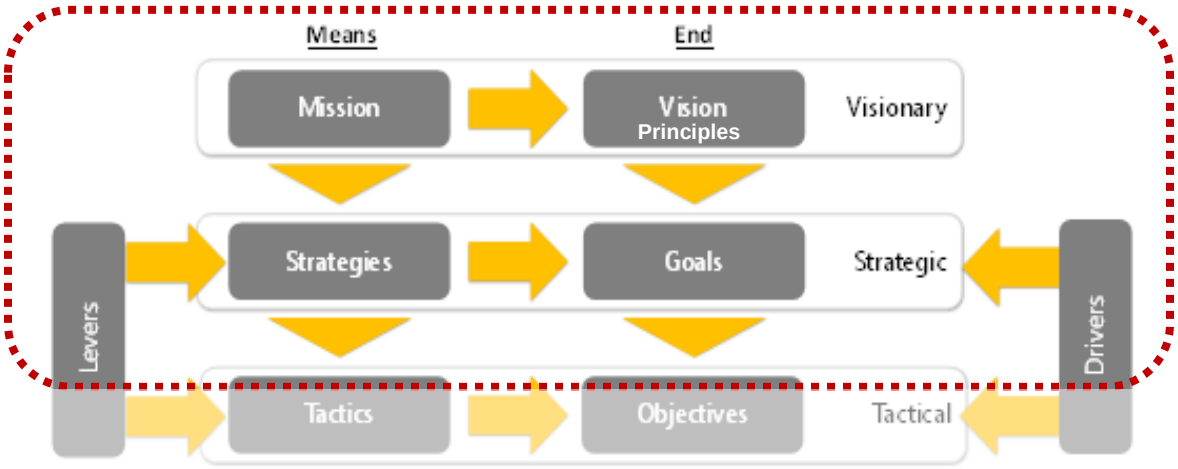
DATA MANAGEMENT STRATEGIES

There are three broad data management strategies that will guide the implementation plan efforts.

1. **Evangelize and implement the Data Governance Framework to the department**
   There needs to be continuing outreach to all levels of the organization to educate about and advocate for the Data Governance Framework. This includes regular reinforcement of principles and goals. It also includes development of policies and standards and adoption of and compliance with those policies and standards. This also includes regular review of principles, goals, objectives, strategies, and tactics to adjust for changing conditions.

2. **Clearly define and implement data governance roles and responsibilities**
   This includes the establishment of the DBHIDS Office of Enterprise Data Management and the identification of the roles within it necessary to ensure success. Data governance requires an enterprise-level focus; it cannot be performed within an operational division. This also includes the definition and adoption of data governance roles and responsibilities for the rest of the organization. Data governance is a partnership between the business and data management, and requires active participation by business management and knowledge workers to be successful.

3. **Manage data as an organizational resource so it is available whenever and wherever needed**
   This is the heart of enterprise data management. It includes the efforts necessary to develop and manage such enterprise resources as the enterprise business model, business glossary, and metadata repository (data architecture). It includes master data domains, reference data, and a data integration and persistence layer (data warehousing). It also includes the capability to provide data whenever and wherever needed to whoever is entitled, with consistent definition and known data quality (business intelligence).



Business Motivation Model (Martin, 2013)

DBHIDS Data Architecture Levers

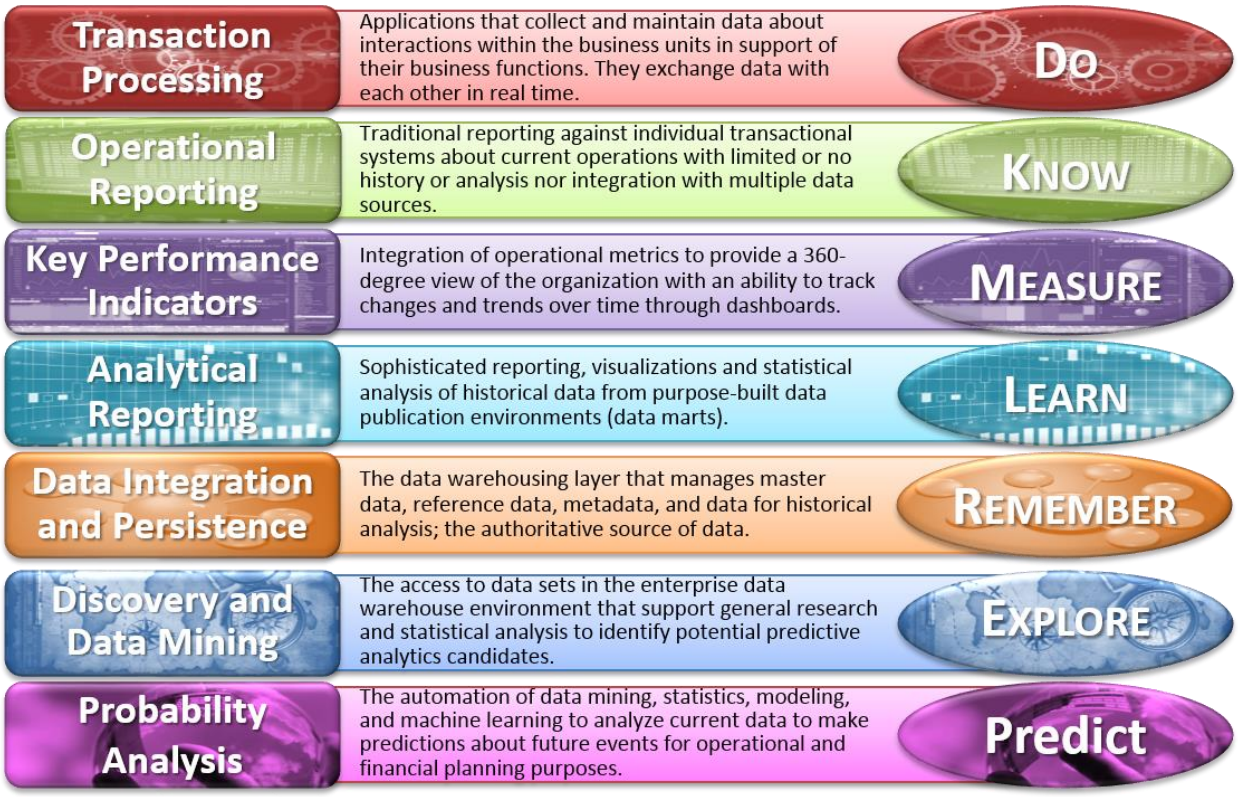
There are two data architecture levers that influence our data governance strategy and implementation. They are the DBHIDS information service delivery use cases and the DBHIDS conceptual information architecture schema. These two levers are the key to achieving our mission and realizing our vision.

DBHIDS Information Service Delivery Use Cases

It is a serious but all too common mistake to apply a one-size-fits-all approach to data management use cases. Organizations attempt to make a system designed for one use case serve the requirements of other use cases. While it is possible, for example, to build a transactional system that also directly supports operational reporting or a data integration layer that also supports analytical reporting, these solutions end up being compromises. (Inmon, Strauss, & Neushloss, 2008)

They are fragile; tightly coupled to processes understood at the time they were built. They lack the agility and flexibility necessary to accommodate new requirements. When they need replacing, the cost is substantially higher due to the unnecessary complexity of the additional overloaded functionality.

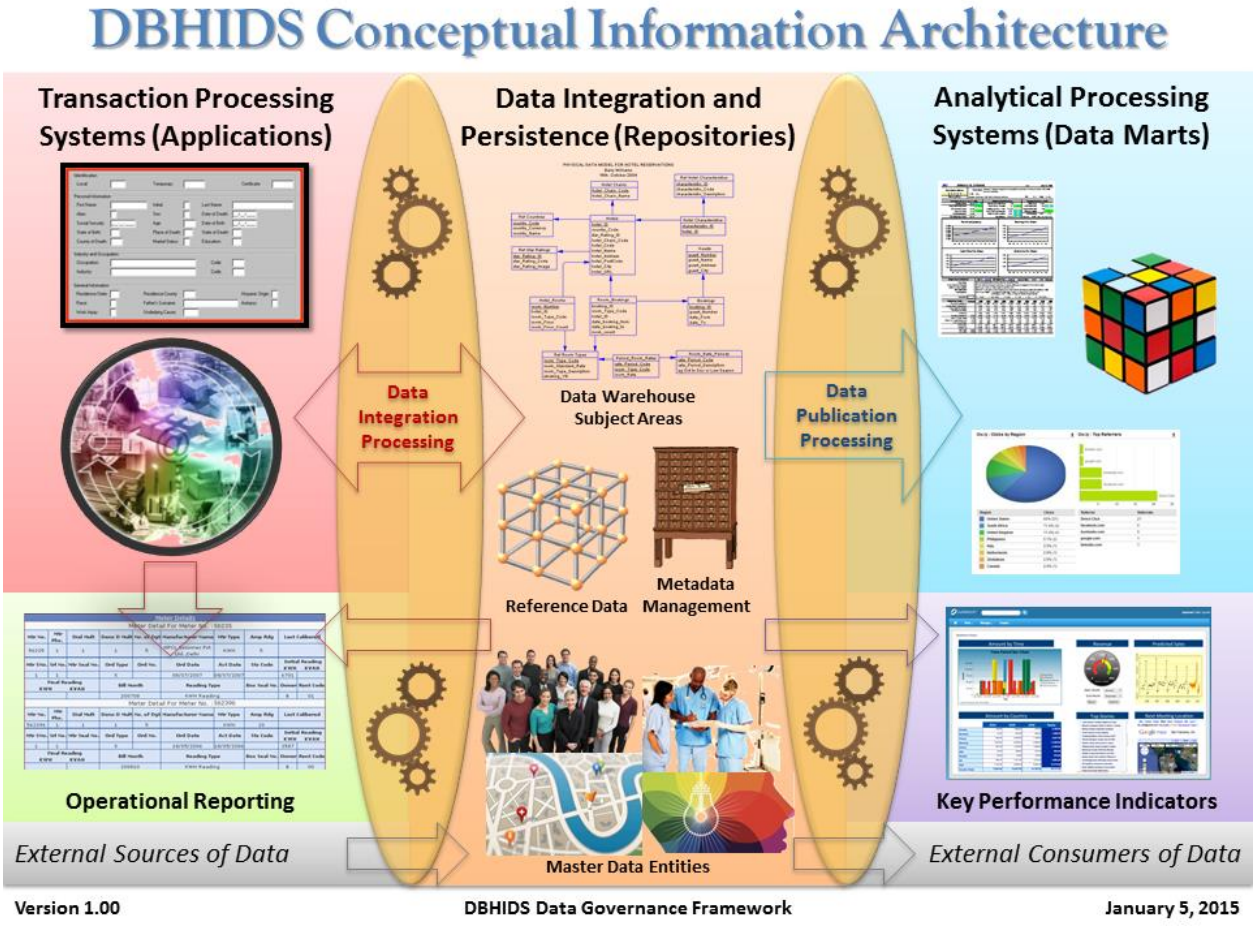
Our seven information service delivery use cases drive the selection of appropriate methodologies and technologies to meet the business needs of the department and provide a single defined source of the truth for business information. The five information service delivery use cases provide the distinctions necessary to guide information management for the department. They categorize the audience and purpose of data so that the proper methodologies and technologies are applied while maintaining a single authoritative source for business data.



Rather than build all functionality into a single solution, it is better to use purpose-built solutions optimized for the desired use case. It is the combincation of a loosely-coupled, service-oriented approach and the data integration and persistence use case that enables all of the other components to function as if members of one unified information ecosystem. (Inmon, Strauss, & Neushloss, 2008)

DBHIDS Conceptual Information Architecture Schema

The following diagram shows the inter-relationships between various conceptual components of the DBHIDS Data Governance Framework. This conceptual model is based in part on the "Corporate Information Factory" as conceived by William H. Inmon, Claudia Imhoff, and Ryan Sousa in the 2000 book of the same name. (Inmon, Imhoff, & Sousa, 2000)



This conceptual model is another form of Data Governance. The value of this model is that it illustrates how all information systems are interrelated conceptually, and that no system should be developed in isolation of existing data, data stores, standards, conventions, or processes. It also reflects the seven information service delivery use cases. DBHIDS solutions are guided by this conceptual architecture.

MOVING FORWARD

For all of the talk about data governance, data architecture, data reusability, and data management, it is important to understand and acknowledge that the purpose of data is first to capture accurately the activity of the business, then help it know what those activities are, then measure the effectiveness of those activities, and finally, to learn from the evaluation of what was measured. In her blog, Beth Kanter describes the mature data-informed organization as one that has reached the "Empowering" stage.

Kanter's Empowering Stage

An organization that has reached the empowering stage…

> *"sets organization-wide key results areas and key performance indicators that are used across programs. Has a staff position responsible for stewarding organization's data, but staff are empowered to check and apply their own data. In addition to weekly check-ins, the organizational dashboard includes key performance metrics related to goals as well as more detailed metrics. The organizational dashboard is shared across departments and there is a process for analyzing, discussing, and applying results. They use data visualization techniques to report the data analysis but also to reflect on best practices culled from the data."* (Kanter, 2011)

These are the notable conclusions that arise from this plan to support achieving data empowerment:

- There is still substantial work needed to create a culture of data governance and to truly implement our data reusability architecture. We must strive to be "Empowered".
- Aligning the vision and mission of this component strategic plan to those of the department will make it easier to convey the value of data governance efforts in meeting department objectives.
- A formal SWOC – Strengths, Weaknesses, Opportunities, and Challenges – analysis provides a baseline for conditions as they existed and for measuring progress towards identified goals. These should be monitored and regular status reports should be produced.
- For DBHIDS to be data-informed will require consistent, dependable, sustainable data of acceptable quality, and this will require involvement of staff throughout the organization. (Ladley, 2012) Once we are data-informed, we can become data-driven.

The following are points of emphasis for each of their respective 2017 goals.

1. The marketing campaign that began in 2015 must continue as business managers have short memories and short attention spans. This must be reinforced by highly visible success stories, no matter how small.
2. A centrally-managed data and information request process must engage the Data Stewardship Council and the Data Governance Executive Board so that they better understand the nature of information requests and the reasons for our data governance strategy.
3. As we begin to deploy self-service BI tools, we must make sure we used role-based security approaches that leverage industry best practices.
4. Once the enterprise data warehouse is in production, we must deliver data marts and dashboards through self-service BI tools that meet eighty percent of business operational and analytical reporting needs.
5. Once the enterprise data warehouse is in production, we must centralize all data feeds and eliminate duplicate data feeds and shadow databases.
6. Master data management must be in place to enable data to be shared; data warehousing must be in place to enable data to be reused.
7. We must complete the acquisition, installation and configuration of the self-service BI platform and begin connecting it to the most requested or useful data sets.
8. We must train current staff to better leverage our enterprise tools and methodologies and recruit new staff with the needed skill sets to grow our data management capabilities.
9. We must complete the enterprise data warehouse and master data management projects and we must acquire the self-service BI reporting platform.
10. We must separate research analysts from data management responsibilities and identify the business analysts and data analysts required by the organization.

Finally, we must remember that the Data Governance Framework Strategic Plan is a domain strategic plan – a strategic plan for a component of the organization. It cannot exist independently of the DBHIDS strategic plan; in fact, it must be consistent with it to be relevant. The purpose of these efforts is not to produce better data, but to help the organization be better informed and make better decisions to fulfill its mission. The organization has made significant progress over the past two years but still has much more to do to realize its data governance vision. We do not anticipate that our principles or use cases will change, but our goals and strategies will evolve to meet new conditions and expectations.

WORKS CITED

Brackett, M. (2011). *Data Resource Simplexity: How Organizations Choose Data Resource Success or Failure.* Bradley Beach: Technics Publishing.

Bryson, J. M. (2011). *Creating Your Strategic Plan.* San Francisco: Jossey-Bass.

Chisholm, M. (2017, July 10). *Fundamental Concepts of Data Quality.* Retrieved July 27, 2017, from First San Francisco Partners: www.firstsanfranciscopartners.com/blog/fundamental-concepts-data-quality

Council for Information Advantage. (2010). *Creating Winning Strategies for Information Advantage.* Hopkinton: EMC Corporation.

Crawford, K., Hasan, H. M., Warne, L., & Linger, H. (2009). From Traditional Knowledge Management in Hierarchical Organizations to a Network Centric Paradigm for a Changing World. *Emergence: Complexity & Organization*, Retrieved from https://emergentpublications.com/ECO/submissions/manuscript_27.pdf.

DalleMule, L., & Davenport, T. H. (2017, May-June). What's Your Data Strategy? *Harvard Business Review*, 112-121.

DAMA International, Inc. (2014). *DAMA-DMBOK2 Framework.* DAMA International, Inc. Retrieved November 28, 2016 from www.dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf.

DBHIDS, City of Philadelphia. (2016). *Data Governance Strategic Plan: Guidelines for Aligning Information Management Concepts, Practices and Context.* Philadelphia: DBHIDS. Retrieved November 21, 2016 from dbhids.org/wp-content/uploads/2016/03/OCIO-DBHIDS-DG-Framework-Strategic-Plan-v1.01.pdf.

Drucker, P. F. (1992). *Managing for the Future: The 1990s and Beyond.* New York: Penguin Books.

Dychê, J. (2008, August 28). *Data Governance in 3-D*. Retrieved June 6, 2017, from BeyeNetwork: http://www.b-eye-network.com/view/8393

Gardner, J. W. (1961). *Excellence: can we be equal and excellent, too?* New York: Harper & Row.

Hay, D. (2011). *UML and Data Modeling: A Reconciliation.* Bradley Beach: Technics.

Inmon, W. H., Imhoff, C., & Sousa, R. (2000). *The Corporate Information Factory.* New York: Wiley.

Inmon, W. H., Strauss, D., & Neushloss, G. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing* . Amsterdam: Elsevier Science.

Kanter, B. (2011, August 17). How Can Nonprofits Switch to a Data-Informed Culture? *www.bethkanter.org/switch-data-driven/*, pp. Retrieved December 3, 2016.

Karl, R. (2008, May 27). *SAP MDM & Data Governance*. Retrieved December 15, 2016, from SAP.com: http://scn.sap.com/blogs/robert.karl/2008/05/27/sap-mdm-data-governance

Ladley, J. (2012). *Data Governance: How to Design, Deploy and Sustain and Effective Data Governance Program.* Waltham: Elsevier.

Loshin, D. (2011). *The Practitioner's Guide to Data Quality Improvement.* Burlington: Elsevier.

Martin, C. (2013, January 14). *The Business Motivation Model – a Key Tool for Talking Business*. Retrieved July 6, 2017, from EA Learning: https://www.ealearning.com/resources/blog/the-business-motivation-model-%E2%80%93-a-key-tool-for-talking-business.html

NASIRE. (1999). *NATIONAL INFORMATION ARCHITECTURE: Toward National Sharing of Governmental Information.* Washington: United States. Bureau of Justice Assistance. Retrieved November 21, 2016 from www.hsdl.org/?view&did=437674.

Seiner, R. (2006, January 1). *The Data Stewardship Approach to Data Governance: Chapter 1*. Retrieved June 6, 2017, from The Data Administration Newsletter: http://tdan.com/the-data-stewardship-approach-to-data-governance-chapter-1/5037

Seiner, R. T. (2005, July 1). Data Steward Roles and Responsibilities. *The Data Administration Newsletter*, pp. Retrieved from http://tdan.com/data-steward-roles-responsibilities/5236.

TechTarget. (n.d.). *Search Data Management*. Retrieved June 6, 2017, from TechTarget Network: http://searchdatamanagement.techtarget.com/definition/data-governance

Thomas, G. (ND). *The DGI Data Governance Framework.* NC: The Data Governance Institute.

Wang, R. Y. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems, Vol. 12, No. 4*, 5-33.

White, W. (2014). *Philadelphia Behavioral Health Services Transformation: Practice Guidelines for Recovery and Resilience Oriented Treatment.* Bloomington: AuthorHouse Publishing.

Wilder-James, E. (2013, December 31). Big Data Variety Means That Metadata Matters. *Forbes*, pp. Retrieved from https://www.forbes.com/sites/edddumbill/2013/12/31/big-data-variety-means-that-metadata-matters/#6310dfdd4c1a.

CREDITS

Vision: *Arthur C. Evans, Ph.D., DBHIDS Commissioner 2004-2017*

Author: Daniel J. Paolini, DBHIDS Chief Information Officer

With substantial contributions from the
DBIDS Enterprise Data Management Team

- Barbara Bunkle
- Minky Kernacs
- Jennifer Li
- Charles Parrish

and input and guidance from the

DBIDS Data Governance Executive Board

DBHIDS Data Stewardship Council

DBHIDS © 2015-2019 – All Rights Reserved

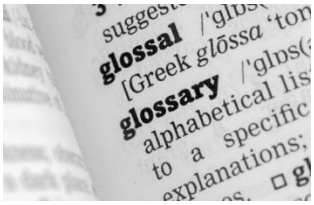
Daniel J. Paolini, CIO – with input from the Enterprise Data Management Team

"I think we may safely say that the studies preliminary to the construction of a great theory should be at least as deliberate and thorough as those that are preliminary to the building of a dwelling-house."

— Charles Sanders Peirce, *The Monist Vol. I, No. 2 (January 1891), p. 161*

APPENDIX A: GLOSSARY OF INFORMATION ARCHITECTURE TERMS

Where possible, the following terms are defined using commonly accepted definitions within the data management profession, but the definitions may be adjusted to represent the way in which these terms are used within the context of the DBHIDS Data Governance Framework and Common Information Architecture.



A B C D E I K L M O P R S T U W

Term	Definition [Home]
ANALYTICAL REPORTING	Analytical reporting is the discovery and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytical reporting relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Analytical reporting often favors data visualization to communicate insight. Analytical reporting is a multidimensional discipline. There is extensive use of mathematics and statistics, the use of descriptive techniques and predictive models to gain valuable knowledge from data and data analysis. The insights from data are used to recommend action or to guide decision making rooted in business context. Thus, analytical reporting is not so much concerned with individual analyses or analysis steps, but with the entire methodology. Since most, though not necessarily all, analytical reporting requires processing to prepare and summarize data the term is often treated as synonymous with online analytical processing (OLAP).
ANALYTICS	Analytics is the discovery, interpretation, and communication of meaningful patterns in data. Analytics is multidisciplinary. Analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Business analytics are the automated processes that standardize the discovery, interpretation, and communication of meaningful data.
ARCHITECTURE	Architecture in the context of information technology is both the process and the product of planning and designing technology-based business solutions and/or their components so as to meet expected requirements and either fit into the existing business and technology environment or improve it in a considered and intended way. See also Data Architecture, Enterprise Architecture, Enterprise Data Architecture, Information Architecture, Service-Oriented Architecture and Solution Architecture.
ATTRIBUTE	An attribute is a specification that defines the logical property of an entity. It may also refer to or set the specific value for a given instance

Term	Definition [Home]
	of such. When instantiated in a physical data model, an attribute becomes a column, field or tag depending on the technology.
BIG DATA	Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, information privacy, search, sharing, storage, transfer, visualization and querying. The term often includes the assumption of the use of predictive analytics or certain other advanced methods to extract value from data, regardless of the size of the data. Big data can be defined by the "V dimensions": Volume, Variety, Velocity, Variability, Veracity, Visualization and Value.
BUSINESS GOAL	A business goal describes a specific outcome an organization expects to achieve over a specific period of time, whereas business objectives are broader and longer in scope. Organizations usually outline their objectives and goals in their strategic plans.
BUSINESS INTELLIGENCE & ANALYTICS UNIT (BIA)	The Business Intelligence & Analytics Unit (BIA) is the unit responsible for determining the information needs of knowledge workers and other data consumers, developing appropriate data reporting structures to meet those needs and providing them to the Data Integration Office, and implementing self-service reporting, analysis and visualization (business intelligence) tools so they are available throughout the organization.
BUSINESS INTELLIGENCE (BI) PUBLISHING TOOLS	Traditionally, business intelligence referred to the tools used to access, query, analyze, or publish integrated data in different types of data warehouses. More and more, the term is being used by vendors to describe the entire data integration and publishing process, which includes the creation of various types of data warehouses. For this reason, our conceptual information architecture refers to the traditional category of tools as BI Publishing tools. These are query, analysis, reporting and visualization tools that provide rapid development of reports and dashboards (graphical representations of key performance indicators) by business people, not technologists. These are the tools that provide meaningful information to data consumers.
BUSINESS MODEL	Business models fall into two general categories, each with two perspectives. The categories are Process Models and Architecture Models. Business process models document the functions and transactions important to an organization and the business rules that guide them. Business architecture models document the information that describes things important to the organization and the business rules that govern that information. Together the models completely

Term	Definition [Home]
	describe the business in a way that can be implemented in a technology solution. Each category of model can be represented in either a conceptual or logical perspective. A conceptual model documents the key processes or data and their basic relationships. A logical model documents the detailed descriptions of the process logic or data attributes.
BUSINESS PROCESS	A business process is a function or transaction performed by a business unit as part of its organizational mission.
BUSINESS RULE	A statement that defines or constrains some aspect of the business as it is implemented in the data model. Data-related business rules are statements, phrased in absolute terms, about data (i.e., a telephone number must have 10 digits) and about relationships between data (i.e., if a phone number is entered, the phone type must be entered).
CHIEF DATA STEWARD (CDS)	Each division or significant administrative unit will designate a Chief Data Steward (CDS). The CDS is a core member of the Data Stewardship Council. The CDS represents the data for which his or her unit is the data stewardship organization. In addition, the CDS works with other data stewards to guide master data management efforts around data sets related to their data needs.
CHIEF INFORMATION OFFICER (CIO)	The DBHIDS Chief Information Officer (CIO) provides strategic vision, executive leadership and coordination for the agency's information management, technology and initiatives.
CHIEF INFORMATION OFFICER, OFFICE OF THE (OCIO)	The Office of the Chief Information Officer (OCIO) is the DBHIDS business unit led by the Chief Information Officer and containing the Office of Enterprise Data Management as well as the individuals responsible for Web 2.0 Coordination, Information Security and Information Technology Planning and Coordination.
CONCEPTUAL DATA MODEL (CDM)	A conceptual data model (also referred to as a Conceptual Schema) is a high-level description of a business's informational needs. It typically includes only the main concepts and the main relationships among them. This is a first-cut, more abstract model, with insufficient detail to build an actual database. A conceptual data model is a map of data concepts and their relationships. This describes the semantics of an organization and represents a series of assertions about its nature. Specifically, it describes the things of significance to an organization (entities/entity-types), about which it is inclined to collect information, and characteristics of (attributes) and associations between pairs of those things of significance (relationships).

Term	Definition [Home]
CONCEPTUAL INFORMATION ARCHITECTURE (CIA)	The DBHIDS Conceptual Information Architecture (CIA) represents the interrelationship of all of the information management components within the department. It categorizes these components by the information service delivery use cases each addresses.
DAMA	The Data Management Association, an organization of professionals working in data architecture, data modeling, data management and related fields.
DATA	A discrete recorded fact or value. Data is the raw material of a system supplied by data producers and used by information producers. Data can be of multiple types, i.e. tabular, spatial, video, audio.
DATA ACQUISITION	The process of obtaining a set of related information about a given subject area. For example, the extraction of source data from a legacy system for the purpose of creating a new data store relating that data to other data (perhaps extracted from another legacy system) is data acquisition. Developing a new application to capture and store information is another form of data acquisition.
DATA AMBASSADOR	The Data Ambassador is the individual within the Data Stewardship Organization identified as the point of contact for questions about the definition and use of data collected or maintained by the DSO within its functional areas and for the documentation of that data in a metadata registry. On behalf of the data stewardship organization, a data ambassador follows and/or implements policies, procedures, and guidelines that pertain to the data during the lifecycle of that data entrusted to his or her stewardship, and participates as a member of the Data Stewardship Council for enterprise data governance issues as well as representing his or her Data Stewardship Organization. The same data may have multiple data ambassadors; e.g. technical, business, collection, publication, etc. The Data Ambassador champions the use of governed data and business intelligence solutions within the business unit.
DATA ANALYTICS	See Analytics.
DATA ARCHITECT (DA)	A Data Architect (DA) is an individual assigned to an IT project, business unit or subject area with the responsibility to develop a logical solution model, a business subject area model or a dimensional model consistent with the enterprise business architecture model. Logical solution models will be used to create physical data models for applications. Business subject area models will guide data integration in the enterprise data warehouse. Dimensional models will define the structures required for reporting and analysis.

Term	Definition	[Home]
DATA ARCHITECTURE	Data Architecture is the orderly arrangement of enterprise resources to achieve: (1) a common understanding of data resources available; (2) a planned approach to data acquisition, storage and retrieval to achieve a high degree of responsiveness to user demands; and (3) a high degree of data sharing and data mobility to reduce program delivery costs.	
DATA ARCHITECTURE UNIT (DAU)	The Data Architecture Unit (DAU) is responsible for maintaining the enterprise business architecture model, reviewing application logical solution models for consistency, overseeing the maintenance of the enterprise business glossary and other metadata resources and for the proper application of enterprise data management policies, standards and practices. The DAU is also responsible for the design of data warehouse subject area models and master entity data and master reference data processes maintained by the Data Integration Office. The DAU is led by the Enterprise Data Architect.	
DATA CONSUMER	A data consumer is an individual or organization that receives data on a regular basis, either in the form of a report or as a data extract file. Typically, this report or data is in support of a predefined, standardized process. A data consumer does not typically make ad hoc one-time requests for data. A data consumer may also be a Knowledge Worker.	
DATA CUSTODIAN	A data custodian is the individual or organization granted authority to possess, use, and/or maintain data in accordance with requirements defined by its data steward. A data custodian is responsible to protect the rights of the data owner for the access, processing, maintenance, storage, protection, and/or destruction of data and electronic records. Data custodians are responsible and accountable for the management and care of the data under their control.	
DATA DICTIONARY	A data dictionary represents information about physical data and database structures. It functions as a catalog of all data objects and elements within a database schema, containing names, structures and information about their usage.	
DATA ECOSYSTEM	See Information Architecture as a deliverable.	
DATA ELEMENT	A data element is an atomic unit of data that has precise meaning or precise semantics. Each data element has a name, a definition and one or more representation terms. Optionally, a data element may have a defined value domain and/or a list of synonyms. Data elements are represented as attributes in a logical model, or as columns, fields or tags in a physical model.	

Term	Definition [Home]
DATA GOVERNANCE	Data governance is the exercise of decision-making and authority for data-related matters. Formal data governance is a system that provides rules and policies proactively to enable ongoing efficient service delivery while providing mechanisms to address data quality issues as they are identified. Data governance also refers to the organizational bodies, rules, decision rights, and accountabilities of people and information systems as they perform information-related processes. In other words, data governance is the identification of those with decision-making responsibility for data management. Data governance is not a technology function. It is driven by the business and forms a bridge between business management and technology providers. Executive sponsorship comes from the business. Data governance and data stewardship represent collaborations of business subject matter experts and information architecture staff. Data management is provided by various technologists overseen by an enterprise information management unit. Where information architecture encompasses data management (the what, when and where), data governance describes roles and responsibilities (the who, how and why).
DATA GOVERNANCE FRAMEWORK (DGF)	The DBHIDS Data Governance Framework (DGF) provides guidance for the department's information architecture and data management activities. It defines the process by which enterprise data is defined and verified as fit for purpose. The DGF also addresses the quality, design, modeling, transport, transformation, storage, access and maintenance of data. It provides a framework for the development and enforcement of policies, standards and guidelines needed to manage data as an enterprise resource with the ultimate goal to deliver effective business intelligence to the user community.
DATA GOVERNANCE UNIT (DGU)	The Data Governance Unit (DGU) is the unit responsible for executing data governance policies and supporting the Data Stewardship Council. The DGU is coterminous with the Office of the Chief Information Officer (OCIO), which provides staff support. All new data initiatives will be submitted to the DGU for review. The Enterprise Data Architect and the DBHIDS IT Planning Coordinator will also participate in the review. The review intensity will be based upon the initiative's risk, size/cost, cross-unit impact and potential to advance department objectives. The review will determine compliance with architecture and standards and identify any need for a waiver.
DATA INTEGRATION	Data integration involves combining data residing in different sources by following agreed upon business rules and definitions to provide

Term	Definition [Home]
	users with a unified view of the data. Data integration becomes more important as the volume and the need to share existing data explodes. Data integration can be performed in bulk with persistent storage using data warehousing technology. Data integration can also be performed in real-time on individual records using service-oriented architecture technology.
DATA INTEGRATION UNIT (DIU)	The Data Integration Unit (DIU) is the unit responsible for integrating and maintaining DBHIDS and external data sources, supplying data to enterprise data marts, business unit data marts and external partners, and managing enterprise master entity data and master reference data. The DIU implements source system integration models provided by the Data Architecture Office and data reporting and analysis models provided by the Business intelligence Office.
DATA INTEGRATION TIER	While all data is an enterprise asset, the governance of that data is the responsibility of multiple business units. To better define data stewardship (governance) responsibilities, the Data Governance Framework classifies data into one of four integration tiers, 0 to 3. See also Universal Data, Enterprise Data, Line-of-Business Data and Programmatic Data.
DATA INTEGRITY	The state that exists when data is handled as intended and is not exposed to accidental or malicious modification, destruction or disclosure. Also, the preservation of data for its intended use.
DATA MANAGEMENT	The technical function of acquiring, defining, certifying, organizing, protecting and delivering data and the metadata that describes it. It focuses technical planning and data operations to meet program delivery objectives and business goals.
DATA MART	A data mart is a data store sourced from a data warehouse that contains smaller subsets of data and focuses on a particular business subject area. In the DBHIDS information architecture it is a pre-defined and pre-formatted subset of data from the enterprise data warehouse (EDW) or an operational data store (ODS) that has been created to address questions that need to be answered by the report community. Data marts are built for the needs of the specific report community, so the same data may exist in many ways and many combinations in different data marts. A data mart may be logical, consisting of views of EDW data, or physical, consisting of extracts of EDW data. Data marts should always receive data from a consistent, integrated source – never directly from individual operational systems – so the answer to the same question from any data mart is always the same.

Term	Definition [Home]
	The DBHIDS Common Information Architecture (CIA) supports the development of dependent data marts (sourced from the EDW environment or an ODS) using conforming dimensions (common reference data used by multiple data marts). When business needs require rapid deployment of a reporting solution, the CIA permits the development of an independent data mart provided that the data to source that mart is first brought into the EDW and staged for future integration.
DATA MODEL	A data model provides a pictorial view of data, groupings of data, relationships between data groupings, and the organization of data groupings by dependencies. There are several perspectives of data models: conceptual, logical, and physical. These models can have different contexts: enterprise, subject area or solution. A conceptual model does not include all of the detailed attributes of an entity. A logical model is fully attributed view that documents both relationships and unique identifiers. A logical model, however, does not reference the characteristics of a particular database system or the physical storage of data. Those would be reflected in a physical model. A solution model is specific to a particular system. A subject area model defines an area of the business without regard to organizational boundaries. An enterprise model serves as a reference that describes data of interest to all or part of an entire enterprise. See also Enterprise Business Architecture Model, Conceptual Data Model, Logical Data Model and Physical Data Model.
DATA OWNER	The data owner is the individual or organization that has statutory rights to the data, no matter who collects the data or who manages it. The rights can include copyright and intellectual property rights as well as the rights to exploit and/or destroy the data. The rights of the data owner apply even when the data is collected by a third party and/or combined with data owned by others. The data owner can be an individual, for data such as personally identifiable information, or an organization, for data such as intellectual property owned by a company. For transaction and derived data collected or created for DBHIDS business purposes by DBHIDS units, the data owner is DBHIDS – not the individual business unit – unless the data owner has been designated as the City of Philadelphia or a state or federal agency.
DATA PERSISTENCE	In computer science, persistence refers to the characteristic of state that outlives the process that created it. Data persistence is the characteristic of data to represent a point in time when it was valid regardless of what changes may have occurred prior or subsequent to

Term	Definition [Home]
	that point in time. Transaction system data is volatile and dynamic so data persistence in best provided in a data warehousing environment.
DATA QUALITY	Data quality is the summary characteristic of data that represents three broad dimensions: 1) fitness for intended use; 2) fidelity to the real-world construct to which it refers; and, 3) internal consistency. Applying ISO 9000:2015, data quality can also be defined as the degree to which a set of characteristics of data fulfills requirements such as: completeness, validity, accuracy, consistency, availability and timeliness. Requirements can be implied or obligatory.
DATA REPOSITORY	Any data container - typically, a database. See also Data Store.
DATA REUSABILITY	Data reusability is the practice of leveraging investments in data cleansing and integration so that future consumers of the data benefit, from initial efforts. Data reusability achieves data integration through standardization of reference and master data and rationalization of definitions and formats for data used across the department. This is more effective than traditional data sharing, where data is moved between units and both the definitions and the transformations of that data are inconsistent.
DATA SECURITY COORDINATOR	The Data Security Coordinator is responsible for identifying data security policies that need to be followed throughout the organization, making implementation recommendations and advising the Data Governance Office.
DATA STAKEHOLDER	A data stakeholder is an individual or organization that is entitled to make use of data collected or maintained within a system or to derive benefits from its collection and/or publication.
DATA STEWARD	A Data Steward is the individual within the Data Stewardship Organization (DSO) identified as the point of contact for questions about the definition and use of data collected or maintained by the DSO within its functional areas and for the documentation of that data in a metadata registry. On behalf of the DSO, a data steward follows and/or implements policies, procedures, and guidelines that pertain to the data during the lifecycle of that data entrusted to his or her stewardship, and participates as a member of the Data Stewardship Council for enterprise data governance issues as well as representing his or her DSO. The data steward establishes business rules, defines data elements, identifies valid data values, establishes certification standards and determines completeness and availability of the data.

Term	Definition [Home]
DATA STEWARDSHIP	Data stewardship is the practice of creating, implementing and following policies, procedures, and guidelines that pertain to specific data during its lifecycle. This includes the establishment of business rules, definition of data elements, identification of valid data values and the establishment of certification standards for the completeness and availability of the data.
DATA STEWARDSHIP ORGANIZATION (DSO)	The data stewardship organization (DSO) is the organization responsible for the definition and use of data collected or maintained within its functional areas. A DSO is responsible for developing policies specifically related to the use of the data, and for designating data stewards to implement and enforce those policies. A DSO may also be the Data Owner or the Data Custodian, but not always (or even usually).
DATA STORE	While the term database is used to indicate just about any formal data storage, it is necessary to categorize databases by their function. The DBHIDS common information architecture provides different roles, both logical and physical, to address these functions. A data store is a database implementation to fulfill one of these specific roles.
DATA WAREHOUSE	Generically, a data warehouse is any informational database, or collection of databases, used to store shareable data for reporting and analysis outside of the system(s) that collected it. The warehouse is usually created through data extracts from operational databases. The warehouse typically allows users to tap into an organization's vast store of operational data to track and respond to business trends and to facilitate forecasting and planning efforts. Multiples types of data stores can all be referred to as "data warehouses", which leads to confusion. The DBHIDS Data Governance Framework does not use the stand-alone term "data warehouse" for that reason. The Gartner IT Glossary differentiates data warehouses from data marts in this way. A data warehouse contains data arranged into abstracted subject areas with time-variant versions of the same records, with an appropriate level of data grain or detail to make it useful across two or more different types of analyses most often deployed with tendencies to third normal form. A data mart contains similarly time-variant and subject-oriented data, but with relationships implying dimensional use of data wherein facts are distinctly separate from dimension data, thus making them more appropriate for single categories of analysis. See also Enterprise Data Warehouse and Data Mart.
DATA WAREHOUSING	Data warehousing encompasses the practices of identifying business data integration requirements, analyzing the source systems that can meet those requirements, modeling data structures to store the

Term	Definition [Home]
	integrated data and designing and building routines to extract, transform and load the source data into a data warehouse structure. See also Enterprise Data Warehousing.
DATABASE	A collection of related data organized to serve one or more applications. In the broader sense, it describes any organized collection of data regardless of the physical storage method. This is often used erroneously as synonym for either the data in the collection or the software product that manages the collection.
DIRECTOR OF BUSINESS INTELLIGENCE & ANALYTICS (DBIA)	The Director of Business Intelligence & Analytics (DBIA) is the individual responsible for supporting a data-driven organization by identifying appropriate subject areas for reporting and evangelizing and implementing ubiquitous business intelligence capabilities that leverage those subject areas.
DIRECTOR OF DATA INTEGRATION (DDI)	The Director of Data Integration (DDI) is the individual responsible for advancing a rational data integration approach within the DBHIDS business and IT communities for the purpose of developing reusable data resources. The DDI shall be responsible for the integrity of data integration processes and for the measurement of data quality against the expected baseline.
DMBOK	DMBOK is the Data Management Book of Knowledge maintained by DAMA, the Data Management Association.
DOMAIN	A domain is a formal category. A data domain refers to all the values which a data element may contain. The rule for determining the domain boundary may be as simple as a data type with an enumerated list of values. An architecture domain refers to one of the three components of enterprise architecture: business process, information or technology. A data management knowledge domain refers to one of the ten knowledge areas defined by the Data Management Association as comprising the Data Management Book of Knowledge (DMBOK) Wheel, in which Data Governance is at the center of nine other knowledge domains.
ENTERPRISE	Generically, the enterprise is that which falls under central direction to meet a common mission. Enterprises are hierarchical and each level of an enterprise, such as a business unit, department or jurisdiction can think of itself as an enterprise while still being a component of a larger enterprise.
ENTERPRISE ARCHITECTURE	Enterprise architecture (EA) is the well-defined practice for conducting enterprise analysis, design, planning, and implementation, using a

Term	Definition [Home]
	holistic approach at all times, for the successful development and execution of strategy. Enterprise architecture applies architecture principles and practices to guide organizations through the business, information, process, and technology changes necessary to execute their strategies. These practices utilize the various aspects of an enterprise to identify, motivate, and achieve these changes. In the context of information technology, enterprise architecture consists of three primary sub-architectures: Business Process, Information and Technology.
ENTERPRISE BUSINESS ARCHITECTURE MODEL (EBAM)	The EBAM is the enterprise logical data model. It defines and standardizes data used to conduct business operations across business units. By documenting the natural relationships between different groups of data, it serves as a starting blueprint for database design activities. The EBAM gives a graphical view of the universal, enterprise and line-of-business data tiers – information that is common to all business units or shared between them. The EBAM supports the management of the overall data assets to achieve optimal integration, sharing, access, and utilization of technology resources and infrastructure. The EBAM is based upon existing and emerging federal reference models and standard industry data models to the greatest extent possible.
ENTERPRISE DATA (TIER 1 DATA)	Enterprise data is federal, state-wide or city-wide data with a common format used by multiple jurisdictions and agencies or to serve common needs, such as geospatial data. It is governed by an external authority, jurisdiction or organization.
ENTERPRISE DATA ARCHITECT (EDA)	The Enterprise Data Architect (EDA) is the individual responsible for advancing the role of data architecture within the DBHIDS business and IT communities. The EDA shall guide the efforts of data architects and analysts within the Office of Enterprise Data Management and the business units in the development of conceptual and logical business models, and logical solution and physical data models. The EDA shall be responsible for overseeing the modeling of master data and the design of master data management processes that will be maintained by the Data Integration Office. The EDA reports to the CIO and is a member of the Data Stewardship Council.
ENTERPRISE DATA GOVERNANCE OFFICER (EDGO)	The Enterprise Data Governance Officer (EDGO) is the individual responsible for ensuring that enterprise data governance policies and practices are followed. The EDGO is responsible for the proper application of the enterprise information and data architecture principals and the overall quality and usability of enterprise data assets. The EDGO leads the Data Stewardship Council.

Term	Definition [Home]
ENTERPRISE DATA MANAGEMENT, OFFICE OF (OEDM)	The Office of Enterprise Data Management (OEDM) within the Office of the CIO is responsible for managing the enterprise data and supporting infrastructure. With guidance from the Data Governance Executive Board and the Data Stewardship Council, OEDM manages this enterprise data environment and conducts efforts specific to enterprise projects or activities.
ENTERPRISE DATA WAREHOUSE (EDW)	From the William Inmon definition, a data warehouse is a purpose-built data store of shareable data in an integrated, subject-oriented, non-volatile and time-variant structure intended to meet operational and analytical requirements to support the organization. From the Bin Jiang, Ph.D. definition, a data warehouse is an infrastructure based on the information technology for an organization to integrate, collect, and prepare data on a regular basis for easing analysis In the DBHIDS common information architecture an enterprise data warehouse adheres to a single enterprise data model to ensure consistency of decision-support data across the enterprise. The DBHIDS EDW is the central integration and storage environment for enterprise data that is gathered from a variety of sources to support data analysis. An enterprise data warehouse is the single version of the truth that supplies historical data to data reusability partners, as well as to analysis areas called data marts. It is not a single database, but a consistent data integration environment that consists of multiple subject areas, staging, archiving and persistent storage and multiple physical databases. It is rarely accessed directly by end-users. The DBHIDS EDW environment can accommodate solutions for a subject area, a business unit or the department as a whole.
ENTERPRISE DATA WAREHOUSING	Enterprise Data Warehousing is the application of data warehousing practices using a centralized approach to create a single source of reusable data for the enterprise.
ENTITY	An entity is a thing capable of an independent existence that can be uniquely identified. An entity is an abstraction from the complexities of a domain. It is some aspect of the real world that can be distinguished from other aspects of the real world. Entities can be thought of as nouns. In data modeling, an entity (as used in entity-relationship modeling) is more accurately an “entity-type”; the inclusive category of all entity instances. An entity, strictly speaking, is an instance of a given entity-type. There are usually many instances of an entity-type. Because the term entity-type is somewhat cumbersome, entity has become an accepted synonym for this term.

Term	Definition [Home]
INFORMATION	A commodity derived from data through analysis or by the orderly presentation of data for human interpretation.
INFORMATION ARCHITECTURE	As a practice, information architecture is the structural design of shared information environments; the art and science of organizing and defining data to support reusability and find-ability. Typically, it involves a model or concept of information that is used and applied to activities which require explicit details of complex information systems. As a deliverable, information architecture represents the formal definition of and mapping between information system components in an enterprise. This indicates which component is responsible for which functionality and the relationship each has with the others. Information architecture can be expressed from a conceptual, logical or physical perspective. The conceptual perspective aligns the information goals of the organization. The logical perspective defines the responsibilities for each component. The physical perspective represents the topology of all of the information systems. See Common Information Architecture.
INFORMATION SECURITY (INFOSEC)	Information security, sometimes shortened to InfoSec, is the practice of defending information from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction. It is a general term that can be used regardless of the form the data may take (e.g. electronic, physical). It encompasses both IT security and information assurance. IT security, sometimes referred to as computer security, is information security applied to technology (most often some form of computer system). In this context, a computer is any device with a processor and some memory. Such devices can range from non-networked standalone devices as simple as calculators, to networked mobile computing devices such as smartphones and tablet computers. Information assurance ensures that data is not lost when critical issues arise. These issues include, but are not limited to: natural disasters, computer/server malfunction, physical theft, or any other instance where data has the potential of being lost.
INFORMATION TECHNOLOGY (IT)	Information technology (IT) is the application of computers and telecommunications equipment to store, retrieve, transmit and manipulate data, often in the context of a business or other enterprise. The term is commonly used as a synonym for computers and computer networks, but it also encompasses other information distribution technologies such as television and telephones.

Term	Definition	[Home]
IT INFRASTRUCTURE	IT infrastructure is shown, as defined in the Information Technology Infrastructure Library (ITIL) standard, version 3, is a combined set of hardware, software, networks, facilities, etc. (including all of the information technology), necessary to develop, test, deliver, monitor, control or support IT services. In practice, infrastructure is used to describe the technology environment within which an application is hosted and/or functions. It can be synonymous with information technology, when information technology is used to describe technologies and not system functionality.	
KEY PERFORMANCE INDICATOR (KPI)	A key performance indicator (KPI) is a type of performance measurement. KPIs evaluate the success of an organization or of a particular activity in which it engages.	
KNOWLEDGE WORKER	A knowledge worker is an individual that uses data to make decisions, measure effectiveness or develop policy. A knowledge worker will assist in identifying the types of reporting and analytic solutions that are required. A knowledge worker may be a subject matter expert but does not have to be.	
LINE-OF-BUSINESS DATA (TIER 2 DATA)	Data that is common to multiple business units in the same line of business, such as Homelessness.	
LOGICAL DATA MODEL (LDM)	A logical data model (LDM) is a tool used to provide a pictorial view of data, groupings of data, relationships between data groupings and the organization of data groupings by dependencies. A logical data model is a view that does not reference the characteristics of a computerized system or of the physical storage of data. An LDM can be for the enterprise, a business subject area or a particular solution. It is a fully attributed view that documents both relationships and unique identifiers. It is created in a fully normalized (non-redundant, logically related) way. It does not reference the characteristics of a particular database system or the physical storage of data. An LDM is derived from and is consistent with the EBAM. It provides documentation of new data structures to the EBAM. An LDM is a prerequisite for any new systems development. It is used to produce a Physical Data Model (PDM). Any changes needed in that PDM should be first captured in the LDM to maintain consistency.	
MASTER DATA	Master data represents the business data objects which are agreed on and shared across the enterprise. Master data can take two forms; either master entity data or master reference data. It can also address two purposes; either operational data quality or analytical reporting.	

Term	Definition [Home]
	Master entity data describes the classes of people, places or things that are of interest to the organization. Examples are Service Recipient, Provider and Location. Master reference data describes the classes of standard code sets that are used throughout the organization. Examples are Program Activity Codes, Diagnostic Codes and Zip Codes. Master data is used to improve operational data quality by providing standard reference data for classification and by integrating standard entities to avoid creating multiple data versions of the same entity. Master Reference Data addresses analytical reporting by providing standardized coding across systems. Master entity data is the glue that connects data from disparate systems.
MASTER DATA DOMAIN (MDD)	A Master Data Domain is entity or reference data of interest to more than one business unit. Most master entity data is managed by multiple master data stewards, whereas most master reference data will have a single master data steward and a single business unit functioning as its data stewardship organization.
MASTER DATA DOMAIN COMMITTEE	The master data domain committee consists of the master data stewards for a specific master data domain. It functions as the data stewardship organization for that domain.
MASTER DATA MANAGEMENT (MDM)	Master Data management (MDM) is comprised of the processes, governance, policies, standards and tools that define and manage the critical entity and reference data of an organization to provide a single point of reference in the organization to ensure consistency and control in the ongoing maintenance, application use and reporting and analysis of this information.
MASTER DATA STEWARD (MDS)	Each master data domain (MDD) will be managed by one or more master data stewards (MDS). A MDS represents the business unit that contributes some or all of the authoritative data in a MDD. Each MDS serves as a core member of the Data Stewardship Council. The MDSs for a MDD comprise the master data domain committee for that domain, serving as its data stewardship organization.
MASTER ENTITY DATA	See Master Data.
METADATA	Metadata is often describes as data about data. It can be conceptual, logical, physical or any combination. Conceptual metadata is in the form of business glossary. Logical metadata is documented in a logical data model. Physical metadata is captured in a data dictionary. Metadata includes attributes such as data name, length, domain of valid values and definition. It also includes business requirements,

Term	Definition [Home]
	business rules, referential integrity, models and other forms of system documentation. Metadata describes how, when and by whom a particular set of data was collected and how the data is formatted. In a data warehouse, Metadata can include elements such as the original source of the data, timestamps, data conversion routines, data transformations, volatility of the data, refresh periods, data reliability indicators and relationships between data from multiple sources.
METADATA MANAGEMENT	Metadata management involves managing data about other data, called metadata, whereas this other data is generally referred to as content data. Forms of metadata include data catalogs, data dictionaries, business glossaries and taxonomies. Metadata management is the end-to-end process and governance framework for creating, controlling, enhancing, attributing, defining and managing a metadata schema, model or other structured aggregation system, either independently or within a repository and the associated supporting processes.
METADATA REPOSITORY	A metadata repository is a database of information describing the characteristics (metadata) of data. Typically, the repository stores a broad range of descriptive information, including business rules, data models and process models that help to elaborate on the usage of data in various systems. Repositories can store metadata for the purpose of identifying and retrieving sets of actual data. Metadata that describes a map is an example.
MODEL-DRIVEN DEVELOPMENT (MDD)	Model-driven development (MDD) is the practice of using conceptual, logical and physical data models to create information system databases consistent with the enterprise definition of data. The DBHIDS information architecture requires a model-driven approach to development. This process begins with the development of a high-level conceptual data model (CDM) to capture the key information needs of the business. After creation of the CDM, the logical data model (LDM) process begins capturing user requirements and business rules to produce a fully normalized and attributed LDM. The LDM is used to produce a physical data model (PDM) for the solution. It is in this PDM that any changes to data structures to address performance, security, or development issues are made. The LDM remains fully normalized and the physical changes are mapped from it. The approved PDM is used to generate the data description language (DDL) needed to create the actual database structures required. When changes later need to be made to the application those changes should first be made in the LDM. The changes are then progressed

Term	Definition [Home]
	through the PDM to the actual database. In this way, the documentation remains accurate and synchronized, and the impact of changes on data integrity is fully understood. In a MDD approach, the conceptual and logical data models are guided by existing overarching models. These models can be in the form of the enterprise business architecture model, subject area models specific to a business area or reference models from external authoritative source.
ONLINE ANALYTICAL PROCESSING (OLAP)	See Analytical Reporting.
ONLINE TRANSACTION PROCESSING (OLTP)	See Transaction Processing.
OPERATIONAL DATA STORE (ODS)	An Operational Data Store (ODS) is a central repository of current operational data gathered from a variety of existing transaction systems to present a single rational view of operational data for a single subject area or business unit, or for an entire agency or line-of-business group. The ODS can serve as a transitional facility while an organization's systems undergo a systematic re-write, as it insulates reports and interfaces from changes in the underlying transaction systems. As legacy systems are re-written, new systems can write directly to the ODS. History should not be stored in the ODS. Reporting can occur directly against an ODS or data can also be replicated into operational reporting areas called operational data marts (opera marts).
OPERATIONAL REPORTING	Operational reporting is reporting about operational details that reflects current activity. Operational reporting is intended to support the day-to-day activities of the organization. Operational reporting is typically performed by the same business unit that is responsible for the transaction processing of the data being reported. Operational reporting is typically produced out of the same transaction processing system or from a copy of it made specifically for reporting purposes. When reporting requires data from other sources, significant transformation of the data or historical data not otherwise required by the transaction processing system, the reporting should be moved into an analytical reporting system.
PHYSICAL DATA MODEL (PDM)	A physical data model (PDM) is what most developers and many business people think of when they hear the term "data model". It may be a relational model or a dimensional model. A PDM is derived

Term	Definition [Home]
	from and is consistent with the LDM. It will document variations from the fully normalized LDM that are necessary for the physical implementation. The PDM is the design of the physical database structures for a system. It is used to produce the physical database. Any changes needed in that database should be first captured in the LDM, and pushed out to the PDM to maintain consistency.
PREDICTIVE ANALYTICS	See Analytics.
PROGRAMMATIC DATA (TIER 3 DATA)	Data owned and managed by a specific business unit for a specific program, and which does not normally need to be shared.
REFERENCE DATA	See Master Data.
REFERENCE DATA MANAGEMENT	See Master Data Management.
REPORTING, ANALYTICAL	See Analytical Reporting.
REPORTING, OPERATIONAL	See Operational Reporting.
SCHEMA	A schema is a description of an information model typically expressed in terms of data structure, domain values, relationships and other constraints. Constraints are expressed using some combination of grammatical rules governing the order of elements, Boolean predicates that the content must satisfy, data types governing the content of elements and attributes, and more specialized rules such as uniqueness and referential integrity constraints. A schema can be expressed in a conceptual, logical or physical perspective. A conceptual schema is represented as a logical data model. A logical schema is represented as a physical data model. A physical schema is represented by the database implementation itself.
SERVICE-ORIENTED ARCHITECTURE (SOA)	A service-oriented architecture (SOA) is an architectural pattern in software design in which application components provide services to other components via a communications protocol, typically over a network. The principles of service-orientation are independent of any vendor, product or technology. As an architectural style it supports service-orientation. Service-orientation is a way of thinking in terms of services and service-based development and the outcomes of services. A service is a self-contained unit of functionality, an operation that may be discretely invoked. A service is a logical representation of a repeatable business activity that has a specified outcome. It is self-

Term	Definition [Home]
	contained. A service may be composed of other services but it will appear as a "black box" to consumers of the service.
SOLUTION ARCHITECTURE	As an activity, solution architecture is the practice of defining and describing the design of a system delivered in context of a specific solution and as such it may encompass description of an entire system or only its specific parts. As a deliverable, solution architecture is the description of a discrete and focused business operation or activity and how information technology supports that operation. An individual solution architecture typically applies to a single project or project release, assisting in the translation of requirements into a solution vision, high-level business and/or IT system specifications, and a portfolio of implementation tasks. Solution architecture activities and deliverables combine guidance from different enterprise architecture viewpoints (business process, information and technical).
SOURCE SYSTEM OF RECORD	A source system of record is the authoritative system that provides data for integration, aggregation, reporting and analysis. It is typically a transaction processing system but may be an internal or external reference system.
SOURCE-TO-TARGET MAPPING	Source-to-target mapping is the process of evaluating each field in a target environment to determine if there is a corresponding field in one or more source environments that is suitable for populating the target field. This process is generally performed during a gap analysis to surface problems resulting from inadequate source system data. The expected result of a source to target mapping exercise is the identification of the authoritative source field(s) to provide the necessary data for each target field.
SPATIAL DATA	Information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. This information may be derived from sources such as remote sensing, mapping and surveying technologies. It includes both attributes (text) as well as spatial (map) data.
SUBJECT AREA	A subject area is the category of knowledge around a specific business function. Subject areas often represent functions that cross organizational boundaries and in that roles become the focus for data integration or master data management efforts.
SUBJECT MATTER EXPERT (SME)	A subject matter expert (SME) is an individual within a business unit or the analyst assigned to the business unit that has the most complete understanding of a particular subject area or process. A SME is engaged to make sure that the right data is being used appropriately

Term	Definition [Home]
	to meet business needs for both transaction processing and reporting purposes. A SME may also be a data steward and/or a knowledge worker, but does not have to be.
TIER 0 DATA	See Universal Data.
TIER 1 DATA	See Enterprise Data.
TIER 2 DATA	See Line-of-Business Data.
TIER 3 DATA	See Programmatic Data.
TRANSACTION	A transaction symbolizes a unit of work performed within a database management system against a database, and treated in a coherent and reliable way independent of other transactions. A transaction generally represents any change in the database. By definition, it must be atomic, consistent, isolated and durable. These properties of database transactions are collectively referred to by the acronym ACID. Transactions have two main purposes to support consistency: 1) to provide reliable units of work that allow proper recovery from failures; and 2) to provide isolation between programs accessing a database concurrently. If this isolation is not provided, the programs' outcomes can be inconsistent and/or inaccurate.
TRANSACTION PROCESSING	Transaction processing is information processing that is divided into individual, indivisible operations called transactions. Each transaction must succeed or fail as a complete unit; it can never be only partially complete. Transaction processing is designed to maintain database integrity in a known, consistent state, by ensuring that interdependent operations on the data are either all completed successfully or all canceled successfully. Since most, though not necessarily all, transaction processing today is interactive the term is often treated as synonymous with online transaction processing (OLTP).
TRANSACTION PROCESSING SYSTEMS	These data stores are where the results of business transactions with the organization are stored. They can be in relational, hierarchical, or file-based database management systems. They can be on a mainframe or on a distributed (network) server. They can be batch processing systems, on-line transaction processing (OLTP) systems, or a hybrid. These systems are typically built to perform a specific business function.

Term	Definition	[Home]
	An organization can have hundreds, or even thousands, of different transaction processing systems. Because many of these systems were built in isolation without concern for future data integration, they are referred to as silo or stovepipe applications.	
UNIVERSAL DATA (TIER 0 DATA)	Common data that is used by most, if not all business units, such as people places and things, but which describes attributes that exist independently of interactions with the business systems, such as an individual's name or date of birth.	
UNSTRUCTURED DATA	Unstructured data refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored either in highly-structured form in databases or annotated (semantically tagged) in documents. The term is somewhat imprecise insofar as this type of data generally has some form of structure, just not the formally defined structure of databases. Examples include emails as well as text documents accompanied by metadata. In this context, unstructured means that whatever structure exists is insufficient for direct, automated processing of the data within the structures. It may also be called semi-structured data or non-tabular data.	
USE CASE	A use case is a list of actions or event steps, typically defining the interactions between a role and a system to achieve a goal. The role can be a human, an external system, or time. In systems engineering, use cases are used at a higher level than within software engineering, often representing missions or stakeholder goals.	
WEB SERVICE	A Web service is a service offered by an electronic device to another electronic device, communicating with each other via the World wide web. In a web service, web technology such as the hyper-text transport protocol (HTTP) is utilized for transferring machine readable file formats such as XML and JSON.	

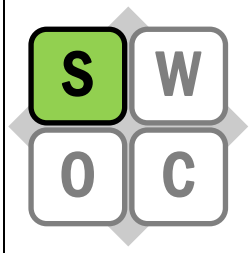
APPENDIX B
LADLEY'S GENERALLY ACCEPTED INFORMATION PRINCIPLES ™

Principle	Description
Content as Asset	Data and content of all types are assets with all the characteristics of any other asset. Therefore, they should be managed, secured and accounted for as other material or financial assets.
Real Value	There is value in all data and content, based on their contribution to an organization's business/operational objectives, their intrinsic marketability, and/or their contribution to the organization's Goodwill (balance sheet) valuation.
Going Concern	Data and content are not viewed as temporary means to achieve results (or merely as a business by-product), but are critical to successful, ongoing business operations and management.
Risk	There is risk associated with data and content. The risk must be formally recognized, either as a liability or through incurring costs to manage and reduce the inherent risk.
Due Diligence	If a risk is known, it must be reported. If a risk is possible, it must be confirmed.
Quality	The relevance, meaning, accuracy, and life cycle of data and content can affect the financial status of an organization.
Audit	The accuracy of data and content is subject to periodic audit by an independent body.
Accountability	An organization must identify parties which are ultimately responsible for data and content assets.
Liability	The risks in information means there is a financial liability inherent in all data or content that is based on regulatory and ethical misuse or management.

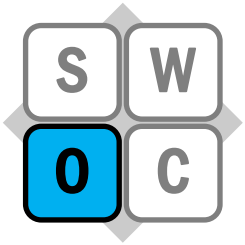
Ladley, J. (2012). *Data Governance: How to Design, Deploy and Sustain and Effective Data Governance Program.* Waltham: Elsevier.

APPENDIX C: STRENGTHS, WEAKNESSES, OPPORTUNITIES, AND CHALLENGES (SWOC) ANALYSIS

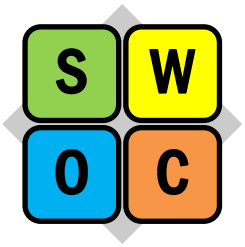
Category	2015 Description	2017 Description
Purpose	The purpose was described passively: *The Data Governance Framework represents the desired future-state needed to address the problems that compromise our information asset today. It documents the Department's Information Architecture. Taking our cue from the Practice Guidelines, we will align information management concepts, practices and context.*	The purpose will be stated more firmly: To be useful, data must be of known quality. To have known quality, it must be governed. Data governance is the responsibility of the entire organization. The Data Governance Framework describes Information Delivery Use Cases, Principles, and Goals that define our data governance strategy as well as the roles and responsibilities to implement it.
Comments	In 2014, the organization knew something was not right but did not know what or how to address it.	In 2017, the organization has a much better understanding of what needs to be done and why. Its difficulty is in dedicating the resources needed to accomplish necessary changes and in breaking old habits.

Category	2015 Description	2017 Description
Strengths 	• A pool of research talent • Initial attempts at data integration in OMH division • Parameterized operational reporting in OMH division • Availability of analytical reporting in CBH division • Commitment by the Commissioner to addressing data issues	• Establishment of Enterprise Data Management unit, including Enterprise Data Architect and Directors of Data Integration and of Business Intelligence • Data Governance Executive Board in place • Data Stewardship Council in place and active • Domain data steward committees established for master data domains of person, provider and service • Enterprise data warehouse effort well underway • Master data management effort well underway • Identification of business intelligence platform to be acquired • Establishment of data analyst coordination with enterprise to reuse data and processes and gain consistency • Commitment by executive management to addressing data issues

Category	2015 Description	2017 Description
Weaknesses 	• No centralized/enterprise data management • No data governance or responsibility for data quality • Most requests for operational data were treated as custom research projects, with data and effort needlessly duplicated while at the same time producing poor data quality • Unwillingness of business unit management to accept responsibility for data governance or prioritization • Unrealistic expectations by executive management that disrupted any possibility of improving data quality by absorbing all available resources for ad hoc data requests • Reliance on outside vendors and contractors for data architecture and design decisions	• Some requests for operational data are still treated as custom research projects, with data and effort needlessly duplicated while at the same time producing poor data quality • Unwillingness of business unit management to accept responsibility for prioritization of the data requests • Some continued reliance on outside vendors and contractors for data architecture and design decisions

Category	2015 Description	2017 Description
Opportunities 	• Enterprise data management is a blank sheet of paper • A number of new initiatives are forcing executive and business unit management to acknowledge the need for data governance, business involvement and enterprise data management • The hiring of a new Chief Information Officer with a data governance, data management and information architecture background meant that the organization would not have to learn the hard way how to address its weaknesses	• Significant turnover in various analyst positions has emphasized the danger of relying on "hero-based" analysts instead of documented and governed data • Several successful small-scale data governance initiatives have demonstrated value to executive and business management • An experience enterprise data architect has been located and hired for a long-term engagement to lead data governance and data integration efforts • A capable Data Integration Director has been identified within existing staff and has assumed those duties • A capable Business Intelligence Director has been identified within existing staff and has assumed those duties • Acquisition of new data and report analysts with a reporting relationship to the Business Intelligence Director to better coordinate activities • Recognition, tolerance and acceptance of the need for data to be managed collaboratively, not competitively • We need to develop a standard data request process that will enable a data governed approach that leverages existing data, tools and methodologies and maximizes the ability to leverage efforts now and in the future

Category	2015 Description	2017 Description
Challenges 	• The organization relied on "hero-based" data management with individual analysts keeping institutional knowledge about data in their heads • Previous attempts at centralized reporting or data management were not successful • The organization had no idea of the value of data architecture and completely lacked any logical understanding of data of interest to the business units • The organization did not have a single data integration area; in fact, each business unit competed to "integrate" data • The organization did not have the ability to deliver self-service reporting, and research analysts took great pride in being identified as the person to whom one had to go to get answers or information • Research analysts, business analysts and IT analysts were in constant competition to see who had "the best" (sic) data • Efforts at normalizing data, rationalizing processes or even creating required documentation were seen as unhelpful, unnecessary and a distraction to getting work done	• The organization is still too "research analyst" top-heavy and must identify or acquire business analysts and BI specialists that can better meet business unit demand for data • The data integration and master data management initiatives currently underway must compete with the day-to-day needs for data and information • Even with substantial marketing over the past two years and several successes, the organization still too easily slips back into the "old way" of approaching data needs • The organization is dependent on the City's central IT organization for many IT services, and its capabilities do not always (or usually) align with our needs and expectations

Category	2015 Description	2017 Description
Conclusions and Recommendations 	In 2015, we identified nine data governance goals. An estimate of their status as of April 2017 is indicated using a 0 (not started) to 10 (completed) scale, shown graphically. 1. Create an information-centric and informed organizational culture. ☺☺☺☺☺☺☹☹☹☹ **2.** Establish a data governance program to provide accountability for information assets. ☺☺☺☺☺☺☺☺☺☺ 3. Provide for effective and appropriate information security. ☺☺☺☺☺☺☺☺☹☹ 4. Improve the quality and usefulness of information by making it timelier, more accurate, more complete and more accessible. ☺☺☺☺☺☹☹☹☹☹ 5. Reduce the costs of managing information. ☺☺☺☺☹☹☹☹☹☹ 6. Share data through reusable processes; reuse data through shared processes. ☺☺☺☺☺☹☹☹☹☹ 7. Provide self-service business intelligence capabilities. ☺☺☹☹☹☹☹☹☹☹ 8. Develop enterprise-class data management staff. ☺☺☺☺☺☺☺☹☹☹ 9. Adopt enterprise-class data management tools. ☺☺☺☺☺☺☺☺☺☹	*Goal 2 was completely achieved and has been reworded to encourage continuous improvement. The other eight goals are retained. The estimated completion status for each goal as of April 2017 is indicated in parenthesis.* 1. Create an information-centric and informed organizational culture. (60%) 2. Continue and strengthen the data governance program by adopting a centrally-managed data and information request process. (40%) 3. Provide for effective and appropriate information security. (80%) 4. Improve the quality and usefulness of information by making it timelier, more accurate, more complete and more accessible. (50%) 5. Reduce the costs of managing information. (40%) 6. Share data through reusable processes; reuse data through shared processes. (50%) 7. Provide self-service business intelligence capabilities. (20%) 8. Develop enterprise-class data management staff. (70%) 9. Adopt enterprise-class data management tools. (90%) *The following goal is added to address the analyst mismatch:* 10. Redefine the roles of data-related analysts in the organization to better match analyst capabilities to organization needs. (30%)

APPENDIX D: DATA STRATEGY ASSESSMENT

The following assessment tool was published in conjunction with an article in the Harvard Business Review by Leandro Dalle Mule and Thomas H. Davenport. (DalleMule & Davenport, 2017) It can be accessed at **hbr.org/2017/05/whats-your-data-strategy**. Items in green represent the eight items that informed the strategy assessment.

Assess Your Strategy Position

To determine where your organization falls on the data-strategy spectrum, select the eight objectives that are most important to your business. (Select only eight.)

- ☐ Create new products and services
- ☑ **Develop analytics and digital capabilities**
- ☐ Generate return on investments in big data and analytics infrastructure
- ☐ Improve IT infrastructure and reduce data-related costs (number of databases, etc.)
- ☐ Improve revenue through cross-sell, pricing, and expanded customer base
- ☑ **Improve the quality of data**
- ☑ **Leverage new sources of data, internal or external**
- ☑ **Meet industry regulatory requirements**
- ☑ **Mitigate operational risks such as data breaks, fraud, etc.**
- ☐ Monetize the value of the company's data; use internal data as a product or service
- ☑ **Optimize existing strong bench of analysts and data scientists**
- ☐ Prevent cyber-attacks and data breaches
- ☑ **Rationalize multiple sources of the same data and information**
- ☑ **Reduce general operating expenses and streamline business processes**
- ☐ Respond rapidly to competitors and market changes
- ☐ Use sophisticated, real-time or near real-time analytics for business

Our results:

> ***Your data strategy is more defensive than offensive****, emphasizing regulatory compliance and data control, security, privacy, integrity, and quality. Organizations in heavily regulated environments where tight data control is critical such as insurance, financial services, and healthcare typically embrace more defensive strategies.*

This strategy is supported by centralized IT managing a Single Source of Truth (SSOT) on behalf of the enterprise, yet making it possible for that SSOT to support multiple presentation contexts, or versions of the truth (MVOT).

APPENDIX E: DATA GOVERNANCE ROLES AND RESPONSIBILITIES

For additional information, see the **MAPPING OF RESPONSIBILITIES TO ROLES** in Appendix F.

Data Governance Bodies and Their Roles and Responsibilities

Ultimately, data governance is a business responsibility. Executive management must endorse it. Senior management must lead it. Involvement by business unit representatives is essential to properly identify, define, approve as fit for purpose, and appropriately use data. While the Data Governance Office and its responsibilities as the facilitator, enabler, and implementer of data governance sits within the Office of Enterprise Data Management, the ultimate responsibility for enterprise data governance resides with management. Enterprise Data Management follows from Enterprise Data Governance. These bodies must be established and become sustainable components

Executive Management Core

The Executive Management Core (EMC) team establishes the business vision, mission, and priorities of the Department. It resolves data governance and data management priorities against other organizational priorities. Led by the Commissioner, it approves data management programs and policies recommended by the Data Governance Executive Board.

Data Governance Executive Board

The Data Governance Executive Board (DGEB) is responsible for approving data governance and data management policies, resolving data governance issues presented by the Data Stewardship Council and prioritizing enterprise data management efforts and initiatives.

The Deputy Commissioner for Administration and Finance leads the DGEB. It consists of all DBHIDS Data Stewards along with members of Executive Management and Enterprise Data Management. It may include others as designated by the Commissioner. It serves as the de facto executive sponsor of enterprise data governance and management initiatives. OEDM staff provides administrative support.

Data Stewardship Council

The Data Stewardship Council (DSC) is responsible for defining enterprise data, developing data management policies, and guiding efforts for enterprise information architecture. Ultimately, its responsibility is to help the Department improve its data quality.

The Enterprise Data Governance Officer or designee leads the DSC. It consists of Enterprise Data Management leaders, the Data Architects, the Master Data Stewards, and Data Stewards from Department business units and systems. It is also open to analysts from the Department's various analytic and IT development units as well as OEDM staff.

Enterprise Data Management Units and Their Roles and Responsibilities

The Office of Enterprise Data Management is the focal point for information architecture and data management for DBHIDS. Reporting to the Director of Enterprise Data Services, it is comprised of units for the enterprise activities of data management, architecture, governance, integration and business intelligence. Each unit has a designated lead employee.

Enterprise Data Governance

Enterprise Data Governance (EDG) is responsible for executing data governance policies and supporting the Data Stewardship Council. EDG is coterminous with the Office of Enterprise Data Services (OEDS), which provides staff support. EDG and IT Strategic Planning review all new data initiatives. It bases the review intensity upon the initiative's risk, size/cost, cross-unit impact, and potential to advance department objectives. The review will determine compliance with architecture and standards and identify any need for a waiver from them.

Enterprise Data Governance Officer

The DBHIDS Director of Enterprise Data Services (DEDS) or designee serves as the Enterprise Data Governance Officer (EDGO) for the Department. In this role, the EDGO is responsible for ensuring that employees follow policies and practices for enterprise data governance. The EDGO is responsible for the proper application of the enterprise information and data architecture principals and the overall quality and usability of enterprise data assets. The EDGO or his or her designee leads the Data Stewardship Council and sits on the Data Governance Executive Board.

Enterprise Data Architecture

Enterprise Data Architecture (EDA) is responsible for maintaining the Enterprise Business Architecture Model, reviewing application logical solution models for consistency, and overseeing the maintenance of the enterprise business glossary and other metadata resources. It is responsible for the proper application of policies, standards and practices for enterprise data management. EDA is also responsible for the design of subject area models in the Enterprise Data Warehouse and for Master Data Management processes maintained by the Enterprise Data Warehousing unit. The Enterprise Data Architect (EDA) leads EDA efforts.

Enterprise Data Architect

The Enterprise Data Architect (EDA) is the individual responsible for advancing the role of data architecture within the DBHIDS business and IT communities. The EDA shall guide the efforts of data architects and analysts within the Office of Enterprise Data Management and the business units in the development of conceptual and logical business models, and logical solution and physical data models. The EDA is responsible for overseeing the modeling of master data and the design of master data management processes maintained by the Enterprise Data Warehousing unit. The EDA reports to the DEDS and is a member of the Data Stewardship Council.

Enterprise Data Warehousing Unit

The Enterprise Data Warehousing Unit (EDWU) is the unit responsible for integrating and maintaining DBHIDS and external data sources, supplying data to enterprise data marts, business unit data marts and external partners, and managing enterprise master entity data and master reference data. The EDWU implements models to integrate source systems as provided by the Data Architecture Unit and models for data reporting and analysis as provided by the Business Intelligence Unit. Its most important role is to provide master data management for the common entities that enable us to connect and integrate data from all systems.

Business Intelligence Unit

The Business Intelligence Unit (BIU) is the unit responsible for determining the information needs of knowledge workers and other data consumers, developing appropriate data reporting structures to meet those needs, and providing them to the Enterprise Data Warehousing Unit, and implementing self-service reporting, analysis, and visualization (business intelligence) tools so they are available throughout the organization.

Other Data Governance Roles and Responsibilities

Data Stewardship Organization

The Data Stewardship Organization (DSO) is the organization responsible for the definition and use of data collected or maintained within its functional areas. A data stewardship organization is responsible for developing policies specifically related to the use of the data, and for designating data stewards to implement and enforce those policies. The Data Stewardship Organization may also be the Data Owner or the Data Custodian, but not always (or even usually). For Master Data domains, the DSO is the Master Data Domain Committee.

Master Data Steward

One or more Master Data Stewards (MDS) shall manage each Master Data Domain (MDD). An MDS is typically a Business Data Analyst. An MDS represents the business unit that contributes some or all of the authoritative data in a MDD. The MDSs are core members of the Data Stewardship Council. The MDSs for a MDD comprise the Master Data Domain Committee for that domain, serving as its Data Stewardship Organization. The MDS is as core member of the Data Stewardship Council.

Chief Data Steward

Each division or significant administrative unit will designate a Chief Data Steward (CDS), typically a Business Data Analyst. Each CDS represents the data for which his or her unit is the data stewardship organization. In addition, the CDS works with other data stewards to guide master data management efforts around data sets related to their data needs. The CDS is a core member of the Data Stewardship Council.

Data Steward

The Data Steward is the individual within the Data Stewardship Organization identified as the point of contact for questions about the definition and use of data collected or maintained by the DSO within its functional areas and for the documentation of that data in a metadata registry. It is often a Business Data Analyst.

On behalf of the data stewardship organization, a Data Steward follows and/or implements policies, procedures, and guidelines that pertain to the data during the lifecycle of that data entrusted to his or her stewardship, and participates as a member of the Data Stewardship Council for data governance issues as well as representing his or her Data Stewardship Organization. The same data may have multiple Data Stewards; e.g. technical, business, collection, publication, etc.

Subject Matter Expert

The Subject Matter Expert is an individual within a business unit or the analyst assigned to the business unit that has the most complete understanding of a particular subject area or process. The Subject Matter Expert is engaged to make sure that consumers use the right data appropriately to meet business needs for both transaction processing and reporting purposes. A Subject Matter Expert may also be a Data Steward or Knowledge Worker, but does not have to be.

Other Data-Specific Roles and Responsibilities

For additional information, see the **OWNER/STEWARD/CUSTODIAN MODEL** in Appendix C.

Data Owner

The Data Owner is the individual or organization that has statutory rights to the data, no matter who collects the data or who manages it. The rights can include copyright and intellectual property rights as well as the rights to exploit and/or destroy the data. The rights of the data owner apply even when a third party collects the data and/or the data is combined with data owned by others.

The Data Owner can be an individual, for data such as personally identifiable information, or an organization, for data such as intellectual property owned by a company. For transactional and derived data collected or created for DBHIDS business purposes by DBHIDS units, the Data Owner is DBHIDS – not the individual business unit – unless an authority designates the Data Owner as the City of Philadelphia or a state or federal agency.

Data Custodian

A Data Custodian is the individual or organization granted authority to possess, use, and/or maintain data in accordance with requirements defined by its Data Steward. A data custodian is responsible to protect the rights of the data owner for the access, processing, maintenance, storage, protection, and/or destruction of data and electronic records. Data custodians are responsible and accountable for the management and care of the data under their control.

Data Stakeholder

The Data Stakeholder is an individual or organization that is entitled to make use of data collected or maintained within a system or to derive benefits from its collection and/or publication.

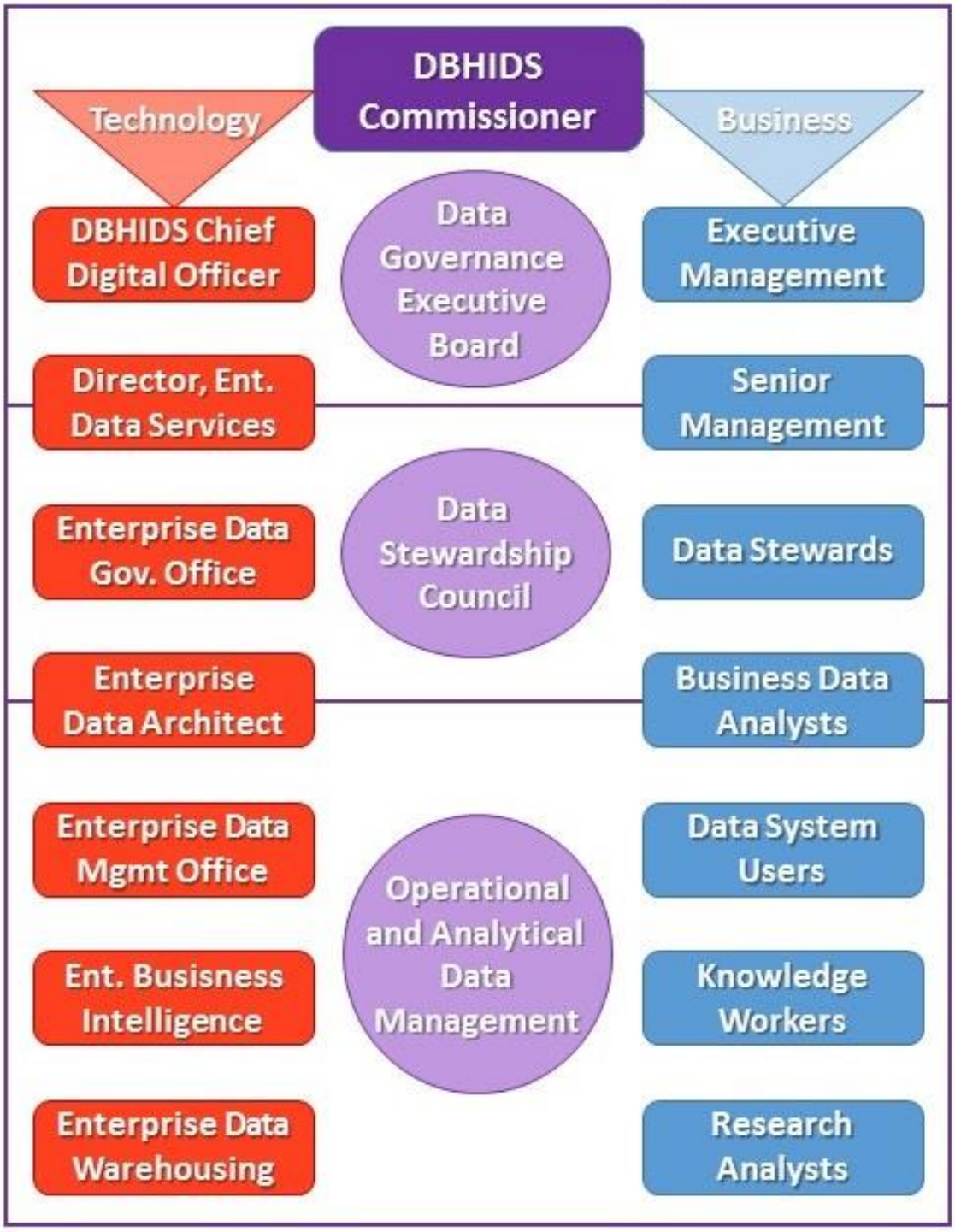
Data Consumer

The Data Consumer is an individual or organization that receives data on a regular basis, in the form of a report or as a data extract file. Typically, this report or data is in support of a predefined, standardized process. A Data Consumer does not typically make ad hoc one-time requests for data. A Data Consumer may be a Knowledge Worker but does not have to be.

Knowledge Worker

The Knowledge Worker is an individual that uses enterprise data to make decisions, measure effectiveness or develop policy. A Knowledge Worker will assist in identifying the types of reporting and analytic solutions that are required. A Knowledge Worker may be a Subject Matter Expert but does not have to be.

Data Governance Roles & Responsibilities



APPENDIX F: THE OWNER/STEWARD/CUSTODIAN MODEL

"Data Owner" is one of the most misused and misunderstood terms in the data arena. It is a component of the Data Owner/Steward/Custodian Model. The relationship between data owner, steward, and custodian is often confusing, as the roles overlap depending on the data. A transaction record can have composite ownership. The transacting organization can own the details of the transaction occurrence while others own the details about people, places, or things involved in the transaction.

The following example illustrates the application of the various roles.

Organization AA must collect the following information from Individual XX to process a transaction:

- Individual XX Name at time of transaction
- Individual XX Date of Birth
- Individual XX Social Security Number
- Individual XX Address at time of transaction
- Transaction Date
- Transaction Amount
- Scanned copy of Individual XX Birth Certificate provided by Organization BB

Because of the transaction, Organization A issues a document with the following information:

- Individual XS Name
- Individual XS Date of Birth
- Individual XS Address at time of transaction
- Document Identifier
- Document Expiration Date

Because of the transaction, as permitted by law or regulation, Organization AA makes the following information available to Organization CC:

- Individual XX Name
- Individual XX Date of Birth
- Individual XX Address at time of transaction
- Document Identifier
- Document XX Expiration Date

The following table indicates who serves in the indicated role for specific data in this transaction.

DATA OWNER/STEWARD/CUSTODIAN MODEL

Data Object	Data Owner	Enterprise Data Stewardship	Business Data Stewardship	Data Custodian
XX Name	Individual XX for the name itself	Organization AA for the transaction, Organization BB for birth certificate	Organization CC for the reused data	Organizations AA, BB, & CC
XX Date of Birth	Individual XX for the date of birth itself	Organization AA for the transaction, Organization BB for birth certificate	Organization CC for the reused data	Organizations AA, BB, & CC
XX Social Security #	Individual XX for the social security number itself	US Social Security Administration	Organization AA for the transaction	Organization AA
XX Address	City of Philadelphia	Master Address Organization	Organization AA for the transaction	Organizations AA & CC
XX Birth Certificate	Individual XX	Organization BB	Organization AA for the transaction	Organization AA
Transaction Amount	DBHIDS	Organization AA	Organization AA	Organizations AA
Document Identifier *	Individual XX or DBHIDS *	Organization AA	Organization AA	Organizations AA & CC
Document Expiration	DBHIDS	Organization AA	Organization AA	Organizations AA & CC
Transaction Record *(this is the transaction as a whole)*	DBHIDS *(subject to personal privacy considerations)*	Organization AA	Organization AA	Organization AA
Transaction Audit Log *(any log created of the transaction)*	DBHIDS *(subject to personal privacy considerations)*	Organization AA	Organization AA	Organization AA
* Where a document identifier, such as a driver's license number or social security number is considered to be personally identifiable information (PII), the owner is the individual to which it refers. Otherwise, it is DBHIDS.				
Note that for DBHIDS-collected data where the data owner is not an external party, the data owner is always DBHIDS, not a specific unit.				