Standard ID:	**IN-DAT-002**
Title:	**Reporting and Data Warehouse**
Domain:	**Information**
Discipline:	**Reporting**
Effective Date: Revision no.: Original date:	**11/10/2011** **0** **11/10/2011**

I. Authority, Applicability and Purpose

A. Authority – Title 29 Chapter 90C Delaware Code, §9004C – General Powers, duties and functions of DTI "2) Create, implement and enforce statewide and agency technology solutions, policies, standards and guidelines, including as recommended by the Technology Investment Council on an ongoing basis and the CIO"

B. Applicability – Applies to all State of Delaware communications and computing resources. DTI is an Executive Branch Agency and has no authority over the customers in Legislative and Judicial Branches, as well as School Districts, and other Federal and Local Government entities that use these resources. However, all users, including these entities, must agree to abide by all policies, standards promulgated by DTI as a condition of funding, access and continued use of these resources.

C. Purpose – Due to the importance of the information managed by the State's technology solutions, it is necessary to establish common guidelines for reporting and warehousing. This document provides approaches and best practices for reporting (static, on-demand and cubes), data marts and data warehouses.

II. Scope

A. State of Delaware – All communications and computing resources involving data, reports, data marts and data warehouses owned by the State of Delaware

B. Areas Covered – This standard covers all aspects of reporting, data marts and data warehouses where data is owned by the State of Delaware.

C. Environments – This standard addresses the tools used to create and manage reports. Also, it addresses the methodologies and practices for creating and managing data marts and data warehouses.

III. Process

A. Adoption – These standards have been adopted by the Department of Technology and Information (DTI) through the Technology and Architecture Standards Committee (TASC) and are applicable to all Information Technology use throughout the state of Delaware.

B. Revision – Technology is constantly evolving; therefore the standards will need to be regularly reviewed. It is the intent of the TASC to review this standard annually. The TASC is open to suggestions and comments from knowledgeable individuals within the state, although we ask that they be channeled through your Information Resource Manager (IRM).

C. Contractors – Contractors or other third parties are required to comply with these standards when proposing technology solutions to DTI or other state entities. Failure to do so could result in rejection by the Delaware Technology Investment Council. For further guidance, or to seek review of a component that is not rated below, contact the TASC at dti_tasc@state.de.us.

D. Implementation responsibility – DTI and/or the organization's technical staff will implement this standard during the course of normal business activities, including business case review, architectural review, project execution and the design, development, or support of systems.

E. Enforcement – DTI will enforce this standard during the course of normal business activities, including business case and architectural review of proposed projects and during the design, development, or support of systems. This standard may also be enforced by others during the course of their normal business activities, including audits and design reviews.

F. Contact us – Any questions or comments should be directed to dti_tasc@state.de.us.

IV. Definitions/Declarations

A. Definitions

1. **Business Intelligence (BI)** is a category of application and technologies for gathering, storing, analyzing and providing access to data to help users with business decisions. Some of the BI tools are OLAP, reporting and data mining.

2. **Data Mart** is a set of data tailored to support specific analytical requirement

   - Customized: for specific reporting purposes
   - Time Variant: The data can report pre and post change information together in a meaningful context
   - Non-Volatile: Once the data is entered into the data mart, it should not change.
   - Historical: The history data required for the data mart will depend on the reporting needs only.
   - Detail: Data is stored at various levels of detail depending on the needs of the business.

- User-friendly presentation: The data mart is customized to the reporting need and helps the user to easily query the data without having to understand the complex level of detail of data warehouse data model.

3. **Data Mining** is a term used to uncover patterns and relationships of data contained within the business activity and history.

4. **Data Provenance** refers to the process of tracing and recording the origins of data and its movement.

5. **Data Warehouse** is a decision support system that typically integrates data from multiple sources into a central location. The integrated data is cleansed, standardized, transformed, and structured in order to more easily analyze the data for a variety of purposes that may cross multiple business areas. A data warehouse is typically used for historical analysis and, therefore, it will contain many years worth of data.
    - Subject Area(s): The data in the data warehouse is organized so that all the data elements relating to the same event or object are linked together.
    - Integrated: The data warehouse contains data from most or all of an organization's operational systems and the data attribute formats are made consistent.
    - Time Variant: The data can report pre and post change information together in a meaningful context.
    - Non-Volatile: Once the data is entered into the data warehouse, it should not change.
    - Detail with much history (Lowest Grain): The data is stored at the lowest level of detail as available in the source system.
    - Enterprise “single source of truth”: Stores all of an organization’s data in a consistent and non-redundant form.

6. **Data Warehouse Architectures**

There are 4 different styles commonly used in the industry are as follows:

- Bill Inmon (Enterprise Approach, Data Warehouse oriented) – The approach of Bill Inmon is called as Top Down approach or Corporate Information Factory (CIF). Bill Inmon saw a need to transfer data from diverse OLTP systems into a centralized place where the data could be used for analysis. He insisted that data should be organized into subject oriented, integrated, non volatile and time variant structures. The data should be accessible at detailed atomic levels by drilling down or at summarized levels by drilling up. The data marts are treated as sub sets of the data warehouse. Each data mart is built for an individual department / business and is optimized for analysis of the particular department / business for which it is created.
- Ralph Kimball (Mart oriented approach) – The approach of Ralph Kimball is called a Bottom-Up approach. Ralph Kimball designed the data warehouse with the data marts connected to it as a bus structure. The bus structure contained all the common elements that are used by data marts such as conformed dimensions, measures etc defined for the enterprise as a whole. He felt that by using these conformed elements, users can query all data marts together. This architecture makes the data warehouse more of a virtual reality than a physical reality. All data marts could be located in one server or could be located on different servers across the enterprise while the data warehouse would be a virtual entity being nothing more than a sum total of all the data marts.
- Hybrid Approach – The Hybrid approach aims to harness the speed and user orientation of the Bottom up approach to the integration of the top-down approach. The Hybrid approach begins with an Entity Relationship diagram of the data marts and a gradual extension of the data marts to extend the enterprise model in a consistent, linear fashion. These data marts are developed using the star schema or dimensional models. The Extract, Transform and Load (ETL) tool is deployed to extract data from the source into a non persistent staging area and then into dimensional data marts that contain both atomic and summary data. The data from the various data marts are then transferred to the data warehouse and query tools are reprogrammed to request summary data from the marts and atomic data from the data warehouse.
- Federated Approach – The approach is referred as a hub and spoke architecture often described as the “architecture of architectures. It recommends an integration of heterogeneous data warehouses, data marts and packaged applications that already exist in the enterprise. The goal is to integrate existing analytic structures wherever possible and to define the “highest value” metrics, dimensions and measures and share and reuse them within existing analytic structures. This may result in the creation of a common staging area to eliminate redundant data feeds or building of a data

warehouse that sources data from multiple data marts, data warehouses or analytic applications.

7. **Online Analytical Processing** (OLAP) is decision support software that allows the user to quickly analyze information that has been summarized into multidimensional views and hierarchies. OLAP tools are used to perform trend analysis. They enable users to drill down into large volumes of data.

   The most common types of OLAP storage are:

   - Relational OLAP (ROLAP) – ROLAP tools extract data from relational databases, using complex SQL statements against relational tables; ROLAP is able to create multidimensional views on the fly. ROLAP tends to be used on data that has a large number of attributes, where it cannot be easily placed into a cube structure.
   - Multidimensional OLAP (MOLAP) – MOLAP tools summarize transactions into multidimensional views ahead of time. Data are organized into a cube structure that can be rotated by the user, which is particularly suited for summaries. Queries are fast because the consolidation has already been done.

OLAP Cube Benefits

OLAP cubes are used for quickly answering analytical queries. OLAP cubes are a part of a broader business intelligence category that also encompasses data mining and relational reporting. Common applications of OLAP cubes include management reporting at State and Agency level, among others. The following are the advantages of OLAP cubes:

- **Quick analysis:** An important advantage of OLAP cubes is that they enhance the speed at which an organization can investigate large data amounts. The OLAP cubes can easily uncover relationships, identify trends and offer numerous perspectives regarding an organization's / State's performance. All these can be attained within a few seconds.
- **Great reporting tool:** OLAP cubes are normally used by an organization for the reporting advantages that they offer. The reporting tool enables the organization to leverage data directly from the cubes and offers scalability for important business intelligence delivered to organizational managers and employees..
- **Improved decision making:** Enables users within an organization to view the analysis from the OLAP cubes, helps in making better decisions and allows the organizations to become more effective.
- **Flexible:** OLAP cubes are flexible and users can easily view their business data using several reporting tools such as Microsoft Excel. They offer enhanced reporting flexibility via highly customized and rich reports.

8. **Reports** are a set of data highly customized for providing business content in a format and type which can be easily used by the user for making better business decisions.

   Reporting Types

   - **Static / Fixed Reporting** - Predefined reports available to users. Report content is predetermined, and can not be modified by users.
   - **Flexible Reporting** - Predefined reports which users can run - have limited control over content by entering parameters for existing elements on the reports. Limited drilling capabilities may exist.
   - **Ad-hoc Query** - Users have the ability to define reports with existing data elements, and control content and layout. Users also have the ability to create metrics, filters, etc.
   - **Executive Information System** - Provides a "dashboard"-style interface to top executives and requires little or no training and data manipulation.

9. **Symmetrical multi processing** (SMP) is the processing of programs by multiple processors that share a common operating system and memory. In symmetric (or "tightly coupled") multiprocessing, the processors share memory and the I/O bus or data path. A single copy of the operating system is in charge of all the processors. SMP, also known as a "shared everything" system, does not usually exceed 16 processors.

B. Declarations

Data Warehouses must:

- Adhere to the State's policies and standards especially the State's Data Management Policy, Data Classification Policy and Delaware Information Security Policy
- Be designed based on the focus of the subject areas of interest with the Enterprise (Agency / State).
  - Enterprise Data Warehouse: Designed based on integrating all the interested subject areas within the State.
  - Department Data Warehouse: Designed based on integrating subject areas of a department, an agency or (2 or more) agencies.
- Be stored at the lowest detail (grain). This allows the data to be aggregated and derived according to business requirements.
- Be built for scalability and operational performance (Able to integrate new subject areas and update existing subject areas in order to respond to business changes).
- Have the requirement for data provenance designed into the subject areas.
- Have documented agreements in place for data sourcing, data processing, data retrieval and disposal.

- Adhere to Federal and State auditability requirements when non-public data is involved.
- Be monitored on a frequent basis for overall data quality. For example, if a street is renamed, the old street name is no longer valid (data decay)
- Be reviewed on a periodic basis (at least, every 2 years) to ensure tha the design and approach still benefit the current processes and are taking advantage of current IT solutions.
- Be regression tested when new subject areas are added or existing subject areas are updated.

Data Marts must:

- Adhere to the State's policies and standards especially the State's Data Management Policy, Data Classification Policy and Delaware Information Security Policy
- Be designed for specific reporting purposes and may contain detail or aggregated data.
- Be derived from a Data Warehouse, if available.
- Be separate from Data Warehouse
- Be reviewed on a periodic basis (at least, every 2 years) to ensure tha the design and approach still benefit the current processes and are taking advantage of current IT solutions.
- Be regression tested when new subject areas are added or existing subject areas are updated.
- Adhere to Federal and State auditability requirements when non-public data is involved.

Data Marts may:

- Have a dimensional schema

Reporting tools must:

- Adhere to the State's policies and standards especially the State's Data Management Policy, Data Classification Policy and Delaware Information Security Policy

Reporting tools may:

- Be used to access data from a Data Mart, which is typically the best approach. When two or more operational data sources are needed in a report, a Data Mart is often created.
- Be used to access data from one operational data source. This is a typical design when the impact to the performance of the operational data source is minimal and the reports are generated in a timely manner. On an annual basis the data custodian should ensure that the operational system and reporting system are performing adequately. Finally, an alternative design would be to backup and restore the database to an alternative location for reporting purposes.
- Be used to access data from Data Warehouse(s) (Enterprise / Departmental)
    - Ensure the Data Warehouse has sufficient resources so that reports (on demand, adhoc and static) cause negligible impact to the Data Warehouse when reports are run against it.
    - Ensure the data retrieval for reporting does not negatively impact the performance of the Data Warehouse and it's down-stream systems.
    - Ensure the security policies (data and user) are enforced based on the nature of data in the warehouse.
    - On a periodic basis (at least, every 2 years) operational / process owner will analyze the warehouse and the reporting systems to ensure that they are still pertinent and performing as per the guidelines.. They will also analyze the impact of any reporting jobs that have been added since the report was created.
    - One recommended approach would be to create reporting tables to be utilized by reporting applications. One implication of this approach is that this will help prevent too many joins or aggregations but may impact other systems.
- Provide the ability to customize the SQL generated by the reporting technology.
- Utilize the role based security provided by the reporting technology
- Have a restart function for Batch reports (Should be able to start from the failed report, so the entire job does not have to be re-run)

Report designs must:

- Adhere to the State's policies and standards especially the State's Data Management Policy, Data Classification Policy and Information Security Policy
- Be able to provide metadata for all terms used in the report. It would be better to have a centrally managed metadata.
- Adhere to Federal and State auditability requirements when non-public data is involved.
- Have the data and time the report is generated.
- Clearly document parameter values chosen by the user, so that this information is available when the report is exported to an external data format or when viewed
- Conform to the State's Data Naming Standard
- Prepopulate data query filters when they are static. (i.e. Do no require the user to continually enter values that are inherent to the report)
- Perform the basic and advanced calculations at the repository or database level.
- When reports are provided online with large volumes of data, release the report details in segments.
- For performance, cache the reports if they will be accessed by multiple parties.

OLAP Cubes must:

- Be customized for specific business / reporting purpose.
- Derived from Data Marts.
- Have proper test cases to ensure the security of the cubes.
- Adhere to Federal and State auditability requirements when non-public data is involved.
- Be a star or snow flake schema.

OLAP Cubes may:

- Have drill down and drill across functionality based on the business requirement.
- Follow these practices for Dimension design for the OLAP tool being used:
  - Do create attribute relationships wherever they exist in the data.
  - Avoid creating attributes that will not be used
  - Do not create redundant attribute relationships
- Follow these practices for Cube design for the OLAP tool being used:
  - Avoid including unrelated measure groups in the same cube
  - Avoid having too many parent-child dimensions in a cube
  - Avoid creating multiple measure groups that have the same dimensionality and granularity
- Follow these practices for Aggregation design for the OLAP tool being used:
  - Do include the granularity attribute of the time dimension in aggregations for measure groups with semi-additive measures
  - Do not build too many aggregations
  - Consider creating separate aggregation designs for partitions with significantly different size or usage

V. Definition of Ratings

Individual components within a Standard will be rated in one of the following categories. **COMPONENT RATING**	**USAGE NOTES**
• **STANDARD** – DTI offers internal support and/or has arranged for external vendor support as well (where applicable). DTI believes the component is robust and solidly positioned in its product life cycle.	These components can be used without explicit DTI approval for both **new projects** and **enhancement** of existing systems.
• **DECLINING** – Deprecated – DTI considers the component to be a likely candidate to have support discontinued in the near future. A deprecated element is one becoming invalid or obsolete.	Via the State's waiver process, these components must be explicitly approved by DTI for **all projects**. They must not be used for **minor enhancement** and **system maintenance** without explicit DTI approval via the State's waiver process.
• **DISALLOWED** – DTI declares the component to be unacceptable for use and will actively intervene to disallow its use when discovered.	No waiver requests for new solutions with this component rating will be considered

A. Applicability of Ratings – The ratings and usage notes are intended to encourage technology decisions to move toward components that enjoy the full support of DTI. However, acknowledging that mass replacement of lower rated components is not feasible, DTI will allow continued maintenance, enhancement, and possibly limited new development using these components. In making such determinations, DTI may require that the requestor demonstrate that they have adequate support arrangements in place.

B. Missing Components – No conclusions should be inferred if a specific component is not listed. Instead, contact the TASC to obtain further information.

VI. Component Assessments

#	Component	Rating	Comments
1	**Reporting Products**		
a)	Cognos/ Report Net	Standard	
b)	Business Objects/Crystal	Standard	
c)	SSAS/SSRS	Standard	
d)	Microsoft Access	Declining	Used as a front-end tool for reporting against back end databases