

New York Health Benefit Exchange

Detailed Design Review Summary for 9.7.4 Data Management Plan October 9 & 10, 2012

<u>Item Number</u>	<u>Topic</u>
9.7.4	Data Management Plan

A red curved line that starts near the top left and arcs towards the right, partially overlapping the text.

New York State Department of Health

New York Health Exchange Project

CSC

Data Management Plan

Document Number: 528500-20120824000
Contract Deliverable Number: P-17
Contract Number: 1106211137
Release Version: 1.0
Date Submitted: August 24, 2012



Version History

Version Number	Implemented By	Revision Date	Approved By	Approval Date	Description of Change
1.0	K. Ernstein, P. Truchard	08/24/2012	C. Adams	08/24/2012	Final for delivery

Table of Contents

1	The NY-HX Data Management Plan Purpose	1
1.1	Scope of the Plan	1
1.2	Plan Organization/Maintenance	1
1.3	Roles and Responsibilities	1
1.4	External Document References	2
2	Data Governance.....	4
2.1	Data Policies, Standards and Procedures	4
2.2	Regulatory Compliance.....	4
2.3	Monitor Conformance with Data Policies, Standards and Procedures	4
3	Data Architecture Management	6
3.1	Develop Exchange Data Model	6
3.2	Define the Database Architecture	6
4	Data Development	6
4.1	Develop Data Modeling and Design Standards	7
4.1.1	Develop Conceptual Data Model	7
4.1.2	Develop Logical Data Model.....	7
4.1.3	Develop Physical Data Model.....	7
4.2	Design Physical Database(s).....	7
5	Data Operations Management.....	8
5.1	Data Recovery Plan.....	8
5.1.1	Failover Environment Planning	8
5.1.2	Disaster Recovery Planning	8
5.1.3	Backup and Recovery Planning	8
5.2	Data Retention	8
5.2.1	Define Data Retention Policies	8
5.2.2	Define Data Archiving Policies	9
5.2.3	Define Data Purge Policies	9
6	Data Security Management.....	10
6.1	Regulatory Requirements.....	10
6.2	Data Access Permissions	10
6.3	Data Masking and Obfuscation.....	10
7	Meta-Data Management.....	11
7.1	Define Meta-Data Requirements	11
7.2	Develop Meta-Data Standards	11
7.3	Create Meta-Data Repository.....	11
8	Data Quality Management	12
8.1	Profile, Analyze and Assess Data Quality	12
8.2	Define Data Quality Metrics	12
8.3	Manage Data Quality Issues	12
Appendix A	GLOSSARY & ACRONYMS.....	13

List of Tables

Table 1- RACI Approach to NY-HX Data Management	2
Table 2- NY-HX Related Documents Reference.....	3
Table 3- Data Conformance Practices.....	4
Table 4- Data Quality Analysis Example	12

1 The NY-HX Data Management Plan Purpose

This document defines how CSC will design and implement Data Management architectures, standards, and governance in support of the New York Health Exchange (NY-HX) project.

1.1 Scope of the Plan

The scope of this plan includes presenting the data governance policies that support all functions of the data domain. All policies and procedures will be defined to meet the regulatory requirements of applicable laws and governing bodies (i.e. HIPAA, HiTech, PCI, etc.). This plan defines the industry accepted principles, including Data Management Body of Knowledge (DM-BOK) and CSC best practices, for enterprise data management that will be used to define the processes for constructing data deliverables for the NY-HX program.

The Data Management Plan serves as the framework for defining policies and procedures that may require further refining and clarification as the program develops, under an Agile project methodology. CSC will work with Department of Health (DOH) staff to assist with further defining items such as data retention, archiving, and data purging.

1.2 Plan Organization/Maintenance

The NY-HX architecture team is responsible for the maintenance of this document and ensures the document is reviewed and updated at least annually. More frequent updates may be made as needed, to accurately reflect changes on the project. Any NY-HX staff member may request a change to this Data Management Plan. This document will be baselined and reside in the program Process Asset Library on the NY-HX SharePoint located at <https://workspace.nyhx.emedny.org/pal/>.

1.3 Roles and Responsibilities

The roles and responsibilities as they pertain to CSC Data Management Plan activities are outlined below. The table below provides a RACI approach to Data Architecture resource involvement.

Table 1- RACI Approach to NY-HX Data Management

Key: R = Responsible; performs the task A = Accountable; ultimately answerable for completion of task C = Consulted; provides information or assistance I = Informed; kept apprised of the activity	Data Architect	Data Specialist	Database Administrator	Meta Data Manager (Database)
Data Architecture Design and Management	RA	R	C	R
Data Modeling	RA	R	C	I
Data Security Model Definition	R	C	RA	I
Meta Data Management	C	C	CI	RA
Data Policy Definition	RA	I	R	C
Data Governance	RA	R	R	R
Data Quality Management	RA	R	R	R
Data Base – Design, Build and Operational Management	R	C	RA	I

The CSC NY-HX architecture team is responsible for the definition of the data architecture and data development activities such as data modeling, data security models and meta-data management. In addition to defining the processes and procedures, the CSC NY-HX team is responsible for creating and maintaining all data models, the operational meta-data repositories, the operational database environment, and implementing data quality systems to enforce defined requirements.

The CSC NY-HX Data Architect is responsible for coordination and overall definition and governance of all data modeling and meta-data management activities. The Data Architect will act as the central point of contact for all data design and data management related tasks.

The Lead CSC NY-HX Database Administrator (DBA) is responsible for the operational database environment including the database engine, database schema and physical database security functions. The Lead DBA will act as the central point of contact for all database management related tasks.

The DOH shall provide staff to support the definition of logical data models, required meta-data information such as business data element definitions, and data quality rule definitions. The DOH shall also approve the data element dictionary and data quality processes.

1.4 External Document References

Reference materials to support the Data Management efforts on this program are found in the following documentation, as shown in the table below.



Table 2- NY-HX Related Documents Reference

NY-HX Facility Management Plan	https://workspace.nyhx.emedny.org/pm/tracker/A-250-00002 att1 NY-HX Facility Management Plan v3.docx
NY-HX System Security Plan	In Process
CSC Standard Operating Procedures for Data Security.	In Process
NY-HX Error and Disaster Recovery Plan	In Process

2 Data Governance

Data Governance, as commonly accepted by Data Management International (DAMA) and affiliates, is defined as the planning, supervision, and control over data management and data use. Below are the key governance areas that CSC uses to define our approach to the overall scope of Data Governance.

2.1 Data Policies, Standards and Procedures

The CSC NY-HX team will define the policies, standards, and procedures related to data that will be used across the NY-HX program. Specific document sets (NY-HX Data Policies and Procedures and NY-HX Data Standards) will be drafted for each policy and maintained in the NY-HX SharePoint repository for all team members to review as required.

Data Polices shall include, but not be limited to:

- Data security to ensure privacy, confidentiality, and appropriate access to all production and non-production data, exchange files, and program information
- Data governance to ensure authority and control (planning, monitoring, and enforcement) of data assets
- Operational data management to ensure proper development, maintenance, and support of structured data including both database management and data technology management
- Data development to ensure consistent, third normal form logical data models, properly physicalized data models, and properly implemented databases
- Meta-data management to ensure that all meta-data meets the overall needs of the Health Exchange including data definitions

2.2 Regulatory Compliance

The CSC NY-HX team will review data requirements against the various laws, rules, and regulations to identify data protection and security requirements. Known sources of these regulations include, but are not limited to:

- The Health Insurance Portability and Accountability Act of 1996 (HIPAA)
- Health Information Technology for Economic and Clinical Health of 2009 (HITECH) Internal Revenue Code section 6103
- Additional Federal Rulemaking with respect to Health Benefits Exchanges
- New York State Information processing policies and procedures

2.3 Monitor Conformance with Data Policies, Standards and Procedures

The CSC approach to data management policy and procedure conformance includes but is not limited to the following practices:

Table 3- Data Conformance Practices

Practice	Benefit
Training	Consistent knowledge across the team of policies, procedures, standards and day to day job practices
Quality Assurance Reviews	The Quality Assurance team monitors data practices and policies, and reviews project deliverables for adherence to project processes and procedures
Third party evaluation	Independent reviews of compliance



State and Federal audits	Continuous monitoring of data conformance using a combination of reports and formal audits consistent with state and federal requirements
--------------------------	---

3 Data Architecture Management

3.1 Develop Exchange Data Model

The NY-HX data model will be made up of three primary components:

- Entity Relationship Diagrams (ERD) which is an abstract and conceptual representation of the data used in the enterprise.
- XML Schema (XSD) which is an abstract collection of metadata consisting of a set of schema components.
- Data Dictionary which is a centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format.

The combination of these components provides the framework to build the three views of the data:

- The conceptual view is the highest level which depicts the data entities but not their relationships.
- The logical view is the next level which includes the relationships between the entities, the cardinality (one to many, one to one), and the beginning stages of normalization; the process of separation of data into key based structures.
- The third level is the normalized physical view.

Additional views may be developed and/or modified for performance reasons in the discovery process when there is actual processing of data.

3.2 Define the Database Architecture

A series of diagrams will be created and maintained that depict the flow of information in and out of the database system. They will depict the input as it is transformed from transactions into the data base engine for placement.

From the design phase to the testing phase, CSC uses an iterative or agile-based methodology and divides the project into multiple releases. Some data definitions will be set up front while others will be progressively defined as the project moves forward.

The data will be stored in a relational format utilizing the LUW DB2 database system. The system will be able to handle data in XML and other formats. The system leverages virtualized platform to maximize performance and high availability (see 5.1.1).

4 Data Development

Data development is the analysis, design, implementation, deployment, and maintenance of data solutions to maximize the value of enterprise data resources. The NY-HX project will utilize an iterative or agile-based methodology and will divide the project into multiple releases during the Requirements, Analysis, and Design sub phases and the Development and Test sub phases. Each release phase is discussed and approved by senior managers from DOH and the CSC team. The Agile methodology will institute a data development process that involves collaboration between management, DOH, data stewards, and subject matter experts in the determination and creation of the data requirements. At specific stages in the development lifecycle, the Data Architect will accept the current version of the data model and the version will be locked down for change management purposes. Any changes proposed to the locked down versions either in development phases or the production phase, will be managed through the NY-HX change control process.

4.1 Develop Data Modeling and Design Standards

IBM Info Sphere Data Architect (IDA) will be used to develop the data model. NY-HX will use Integration Definition for Information Modeling (IDEF1X) notation diagrams to visually represent and develop artifacts of database applications in a single, tightly integrated development environment.

The Data Development process will define a set of data specifications and related diagrams that reflect data requirements and designs. A data model will be developed using IDEF1X standard symbols and text to represent data elements and relationships between them.

4.1.1 Develop Conceptual Data Model

A conceptual data model will be developed to provide a high level visual perspective of each subject area of the NY-HX solution. The subject areas within NY-HX are Plan Management, Eligibility and Enrollment, Financial Management and Customer Communications and Services. It will contain the most basic and critical business entities within each subject area and functions, with a description of each entity and relationships between entities. This data model will take into account information on business entities contained in the CMS Exchange Blueprints, and updated with New York state specific information.

4.1.2 Develop Logical Data Model

The NY-HX system is developed using an Agile/Scrum system development life cycle (SDLC), that will include progressive creation of a Logical Data Model (LDM), to be created and updated by the development team. Entity definition, relationship identification, attribute definition, and data type development will be supported by a DBA embedded within the Scrum development team. Oversight, consistency, and completeness reviews will be conducted by the NY-HX Data Architecture team.

4.1.3 Develop Physical Data Model

IBM Info Sphere Data Architect (IDA) will be used to transform the LDM into a Physical Data Model (PDM) initially. IDA will also be used to maintain and update the LDM and PDM.

4.2 Design Physical Database(s)

IDA will be used to generate Data Definition Language (DDL) statements which can then be deployed to the DB2 database server to create tables, columns and indexes.

5 Data Operations Management

5.1 Data Recovery Plan

The NY-HX system requires a macro and micro level recovery system. The macro level involves recovery of data assets at the server level and data center level. The micro level involves recovery of data with point in time backups on a file or table level. The Error and Disaster Recovery Plan also addresses this topic.

5.1.1 Failover Environment Planning

The system will be architected for high availability utilizing the DB2 HADR feature. This will utilize active and passive database nodes. As one area fails, the second (passive node) will become the primary and remain until the original is deemed usable as outlined in the Error and Disaster Recovery Plan. This process will be defined and supported by a detailed data failover and recovery process that will be created as part of the overall data continuity planning.

5.1.2 Disaster Recovery Planning

The NY-HX Disaster Plan will leverage the eMedNY disaster recovery framework. A disaster recovery node will be maintained and kept in sync at the SunGard disaster recovery site in New Jersey. This process will be defined and supported by a detailed data services failover and recovery process that will be created as part of the overall data continuity planning. The Error and Disaster Recovery Plan explains this process in more detail.

5.1.3 Backup and Recovery Planning

Backups of both data and logs will be run and maintained on a daily, weekly, and monthly cycle. Netback up will provide a "hot" backup of DB2 for the data store. The Veeam product will be utilized for backing up the structural component. These disk backups are transferred to tape and stored off-site per New York State requirements using the same staff, equipment, and strategy in use by eMedNY, the New York Medicaid Management Information System (MMIS).

5.2 Data Retention

This section outlines how NY-HX will retain, archive, and purge data. These policies will be interpreted into specific processes that will be applied to the NY-HX data repositories. It is envisioned that these processes will be automated where it is practical to do so and that the function and outcome of these policies will be reviewed on a regular basis as part of the service management function.

5.2.1 Define Data Retention Policies

The data retention policy will be determined based on State and Federal requirements. After these requirements are met, the appropriate data will be either archived or purged through scheduled, automated processes. The archived data will be retained in a hierarchy that keeps the newest (most active) on a disk system for easy retrieval followed by data retained in a tape format. A regular audit will be conducted to ensure that archive/purge is occurring according to policy and that those policies meet current State and Federal requirements.

5.2.2 Define Data Archiving Policies

User data will be archived based on the defined requirements. System data, including logs will be kept in accordance with the Error and Disaster Recovery Plan. This system data archiving policy will follow existing eMedNY policies to meet State or Federal requirements, unless indicated otherwise.

5.2.3 Define Data Purge Policies

A data purge policy will be developed, documented, and executed on a regular pre-determined basis through scheduled, automated processes. The policy will include system temporary queues with an immediate or very short life span and user data that would be purged on a longer but still agreed upon schedule. The policy will be developed in accordance with the NY-HX solution.

6 Data Security Management

Data security will be managed in accordance with the safeguards identified in the NY-HX Facility Management Plan, the NY-HX System Security Plan, the NY-HX Error and Disaster Recovery Plan, and CSC's standard operating procedures for data security.

6.1 Regulatory Requirements

Data Security provisions of the HIPAA and HITECH Acts, IRS publication 1075, and specific New York State regulations will govern NY-HX data security. All of these frameworks are patterned after NIST Special Publication 800-53A standards. The NY-HX System Security Plan provides more information on specific safeguards.

6.2 Data Access Permissions

DOH, CSC or any other user will be given data access permissions based on database controlled tables. These tables are synchronized with active directory groups. This enables permissions to be enforced across all NY-HX applications. These permissions cover create, read, update, and delete functions. These permissions will be assigned to the various roles that will be engaged in the development, testing and operation of the NY-HX platform. Assignment of these roles to specific users as well as any change to the permissions of roles will be handled via the standard CSC security request and approval process.

6.3 Data Masking and Obfuscation

Protected Health Information (PHI) data and/or Federal Tax Information (FTI) is not expected to exist in non- production environments. Only test data is expected to exist in those environments, however, de-identification processes are available and will be used as needed as a safety measure to help prevent breach. Sensitive data, such as names, account information, medical details are obscured using data masking so they will be untraceable and unusable. The data masking algorithms are applied using a tool across multiple tables, applications, and environments by database administrators so referential and business integrity will always be maintained. Anonymization and pseudonomization procedures, where most identifying fields within a data record are replaced by one or more artificial identifiers, will be implemented on production data.

7 Meta-Data Management

Meta-data is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage data. Meta-data Management is the set of processes that ensure proper creation, storage, integration, and control to support associated usage of meta-data. Meta-data will have all the physical information on tables and attributes defined in the NY-HX database, and will be maintained in an iterative fashion, as part of the Agile methodology. As new meta-data information is discovered throughout the sprint cycles, the meta-data dictionary will be updated and new information added. The NY-HX solution includes a Meta-data management tool to assist in this process.

7.1 Define Meta-Data Requirements

Meta-data includes items such as application code, physical servers, database structures, etc. All have meta-data associated with them, which will be collected throughout the Agile sprint cycles, and will be documented within the data repository. Meta-Data Requirements are gathered from both the business and technical users, confirming the need for a meta-data management environment, to set scope and priorities, educate and communicate. These requirements guide tool evaluation, implementation, modeling, internal standards, and provided services that rely on meta-data to estimate and justify staffing needs.

7.2 Develop Meta-Data Standards

NY-HX Meta-data Standards will be developed using common industry standards such as Universal Modeling Language (UML), Common Warehouse Modeling (CWM), the National Information Exchange Model (NIEM), and more specific healthcare domain standards such as HIPAA and Health Level Seven (HL7). This metadata will be managed by Adaptive Metadata Manager.

7.3 Create Meta-Data Repository

Meta-data Repository will be created initially in accordance with industry standard objects (classes, methods, data element dictionary, file definitions, etc) to be populated during the agile design and development sessions. Additional meta-data objects will be identified and defined throughout the Agile iterative process; those items will be added to the meta-data repository. Adaptive Metadata Manager, a web-based repository, will be used to create the Meta-data Repository. Meta-data will be maintained on a regularly scheduled basis.

8 Data Quality Management

This section defines quality management techniques to measure, assess, improve and ensure the fitness of data for use.

8.1 Profile, Analyze and Assess Data Quality

The Profile, Analyze and Assess Process is an ongoing operational process that consists initially of looking at the actual data. Two key questions that are asked, are whether the data is complete, and whether the data is accurate. Data profiling, analysis and assessment is the process of gaining an understanding of the existing data relative to the quality specifications, as shown below. This is not a complete list but just a sample of what is involved in this process.

Table 4- Data Quality Analysis Example

Issue	Example
Out of Acceptable Range Patient	Age = 185 ()
Non-Standard Data	Main Str, Main Street, Main ST, Main St. 7
Invalid Values	Data can be "A" or "B" but Value = "C"
Differing Cultural Rules	Date = Jan 1, 2002 or 1-1-2002 or 1 Jan 02
Varying Formats	(919)674-2153 or [919]6742153 or 9196742153
Cosmetic	jon j jones transformed into Jon J Jones
Verification	ZIP code does not correspond to correct City & State

NY-HX will use IBM Info Sphere Information Analyzer and Quality Stage to profile, analyze, and assess data quality. Info Sphere Information Analyzer helps understand data quickly by offering data quality assessment, flexible data rules, design, analysis, and quality monitoring capabilities. These insights help derive more information from enterprise data to accelerate information-centric projects. Tight integration with IBM Info Sphere Quality Stage provides direct re-use of data rule definitions created in Information Analyzer.

8.2 Define Data Quality Metrics

Data Quality Metrics will be defined for critical data elements. The CSC NY-HX Team will work, through an iterative process, with the Department of Health to define what the quality metrics are based on profile information collected with each data feed. Data expectations will be defined for each of the data elements. For each data expectation, the associated dimension of data quality and one or more business rules to use to determine conformance of data to expectations will be specified. The process for measuring conformance and acceptability will be specified and be documented in the Info sphere Quality Stage toolset.

8.3 Manage Data Quality Issues

A Data Quality Incident Reporting system will be identified and used to support the enforcement of the data quality reporting. This includes tracking data quality incidents, activities for researching and resolving data quality issues, cleansing, and correcting data quality defects.



Appendix A GLOSSARY & ACRONYMS

Refer to the Master Project Management Deliverables Glossary and Acronym List posted in the Deliverables Section of the NY-HX SharePoint site.