

# New York Health Benefit Exchange

## Blueprint Summary for 9.7.4 Data Management Plan October 26, 2012

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

<b>Version Number</b>	<b>Modified By</b>	<b>Revision Date</b>	<b>Description of Change</b>
1.0	K. Ernstein, P. Truchard	08/24/2012	Original Delivery
1.1	Jerry Daly	10/24/2012	Updates as requested by CCIIO from CMS Design Review on 10/9/2012 and 10/10/2012

---

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

# **New York State Department of Health**

## ***NY-HX Project***

**CSC**

### **Data Management Plan**

Release PRELIMINARY  
Date Submitted: 10/24/2012

## Table of Contents

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

# 1 The NY-HX Data Management Plan

This document represents how the New York Health Exchange (NY-HX) team plans to define and implement Data Management architectures, standards and governance in accordance with standardized practices.

## 1.1 Scope of the Plan

This plan covers applicable, industry accepted data management principles for enterprise data management. Using the Data Management Body of Knowledge (DM-BOK) and CSC best practices, CSC is defining the processes for constructing data deliverables for the NY-HX program.

This document serves as a framework for further definition of policies and procedures which require refining and clarification as the program develops. As an example, CSC's framework for items such as data retention will be flexible in nature to allow for varying data retention, archiving and purge however it will require Department of Health (DOH) staff to help define the exact rules.

## 1.2 NY-HX Data Management Mission and Goals

The mission of the CSC NY-HX team is to define the policies, procedures and guidelines to support the significant data aspects of the Health Exchange. The goal of this team is to create data governance policies that support all functions of the data domain from data modeling to data archiving, retention and purge. All policies and procedures will be defined to meet the regulatory requirements of all applicable laws and governing bodies (i.e. HIPAA, HiTech, PCI, etc.).

## 1.3 Functions and Activities

The CSC NY-HX team is responsible for the definition of the Data Architecture, 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 DOH shall provide staff to support the definition of logical data models; required meta-data information, such as business data element definitions; data quality rule definitions. In addition, the DOH shall approve the data element dictionary and data quality processes.

# 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.

## 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. These documents will be created and commonly published for all NY-HX team members to review as required.

Data Policies 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, 3<sup>rd</sup> 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 Exchange including data definitions.

## 2.2 Regulatory Compliance

The CSC NY-HX team will review all data requirements against the various laws, rules and regulations to ensure that all data is protected and secured in accordance to policy. Some compliance information is documented in other documents, such as the Technical Controls document and the Systems Security Management Plan document. In addition we will describe our compliance with the following laws, acts and regulations that will be enforced:

- HIPAA
- HiTech
- FIPS
- NIST
- Others TBD

## 2.3 Monitor Conformance with Data Policies, Standards and Procedures

Speak with AJ and Rafael Sosa = Need to add what we are going to do here!

# 3 Data Architecture Management

## 3.1 Develop Exchange Data Model

The NY-HX data model will be made up three primary components comprising the Entity Relationship Diagrams (ERD), XML Schema (XSD) and the Data Dictionary. The combination of the three components provides the framework to build the three views of the data. The conceptual view is the highest level depicting the data entities but not their relationships. The second is the logical view 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) , and then finally the normalized physical view. Additional views may be developed and/or modified for performance reasons during the discovery process during actual processing of data.

Framework Components described

- Entity Relationship Diagrams (ERD) is an abstract and conceptual representation of the data used in the enterprise.
- XML Schema (XSD) is an abstract collection of metadata, consisting of a set of schema components.

- Data Dictionary is a centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format.

## 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. They will depict the input as it is transformed from transactions into the data base engine for placement.

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 data resources to the enterprise. NY-HX agile methodology will institute a data development process involving Business Data Stewards, SMEs, Data Architects, Analysts and Database Administrators, Software Architects and Developers in the determination and creation of the data requirements

### 4.1 Develop Data Modeling and Design Standards

IBM InfoSphere 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.

#### 4.1.2 Develop Logical Data Model

A Logical Data Model (LDM) will be developed by the Scrum/Sprints/JAD team through NY-HX Agile process sessions. The Scrum/Sprints/JAD session involves meeting with the client (DOH), System For Electronic Rate and Form Filing (SERFF), Provider Network Data Service (PNDS), Quality Assurance Reporting Requirements (QARR), Welfare Management System (WMS), Children's Health Insurance Program (CHIP) and other external agencies to define/understand the data and develop the LDM.

### **4.1.3 Develop Physical Data Model**

IBM InfoSphere Data Architect (IDA) will be used to transform the Logical Data Model into a Physical Data Model initially. IDA will also be used to maintain/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.

### **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. 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.

### **5.1.3 Backup and Recovery Planning**

A cycle of Daily, Weekly and Monthly backups of both Data and logs will be run and maintained. The Data store will be backed up using Netback up database backup agent, which enables a "hot" backup of DB2. The structural component will be backed up utilizing the Veeam product. This strategy will leverage the backup components currently in use by eMedNY.

## **5.2 Data Retention**

### **5.2.1 Define Data Retention Policies**

The data retention policy will be determined based on State and Federal requirements. After the requirements are met the appropriate data will be either archived or purged. The archived data will be retained in a hierarchy that will keep the newest (most active) on a disk system for easy retrieval followed by data retained in a tape format.

### **5.2.2 Define Data Archiving Policies**

User data will be archived based on the defined requirements. System data including but not limited to logs, will be kept in accordance with the associated system recovery plan. This system data archiving policy would follow standard CSC policies.

### **5.2.3 Define Data Purge Policies**

A data purge policy will be developed documented and executed on a determined regular basis. 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 management is a way to maintain the integrity of data and to make sure that the data is not accessible by unauthorized parties or susceptible to corruption. Data security is put in place to ensure privacy in addition or protecting the data. The secure management of the data will be defined by the various rules and compliance requirements as set out in the security policy and management plans.

### **6.1 Regulatory Requirements**

All HIPAA, State and Federal (Federal Information Security Management Act (FISMA)& IRS Publication 1075 Tax Information Security Guidelines ) requirements will be followed throughout the Security implementation.

### **6.2 Data Access Permissions**

State, 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, as well as tracking and auditing of this data where appropriate. 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 system integrator security request and approval process.

### **6.3 Data Masking and Obfuscation**

Protected Health Information (PHI) and/or Federal Tax Information (FTI) data for all non production environments will be de-identified as a safety measure to help prevent breach. Using data masking, sensitive data - such as names, account information, medical details and so on, are obscured so they will be untraceable and unusable. The data masking algorithms are applied across multiple tables, applications and environments by Database administrators using a Data Masking tool so referential and business integrity will always be maintained. Anonymization and Pseudonymisation procedures where most identifying fields within a data record are replaced by one or more artificial identifiers will be implemented on all appropriate data. Note: No masking in DDI (Design, Development and Implementation) Phase as only test data is expected. Guidance regarding targets of data obfuscation will be sought from all of the above mentioned regulatory requirements in section 6.1.

## **7 Meta-Data Management**

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage data. Metadata is often called data about data or information about information. 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. Meta-data will be maintained as part of NY-HX agile methodology to support the efficient use and control of the data as it is required by the systems.

### **7.1 Define Meta-Data Requirements**

Meta-Data Requirements are gathered to confirm the need for Meta-Data Management environment, to set scope and priorities, educate and communicate, to guide tool evaluation and implementation, guide Meta-Data Modeling, guide internal standards, guide provided services that relay on Meta-Data and to estimate and justify staffing needs. These requirements are obtained from both Business and Technical users in the scrum team during JAD sessions.

### **7.2 Develop Meta-Data Standards**

NY-HX Meta-data Standards will be developed using Industry Standards and will be managed by Adaptive Metadata Manager.

### **7.3 Create Meta-Data Repository**

Meta-data Repository will be created as per the requirements derived from JAD sessions. Adaptive Metadata Manager, a web-based repository, will be used to create the Meta-data Repository. Meta -data will be maintained on a regular scheduled time period.

## 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

Profile, Analyze and Assess process consists initially of looking at the actual data. Two key questions to ask 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.

<b>ISSUE</b>	<b>EXAMPLE</b>
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

HX-NY will use IBM InfoSphere Information Analyzer and QualityStage to profile, analyse and assess data quality. InfoSphere Information Analyzer helps understand data quickly by offering data quality assessment, flexible data rules design & analysis, & quality monitoring capabilities. These insights help derive more information from enterprise data to accelerate information-centric projects. Tight integration with IBM InfoSphere QualityStage 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. 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.

### 8.3 Manage Data Quality Issues

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