Enterprise Data Warehouse Standard Operating Procedures

1. Introduction

1.1 Background

In the course of providing patient care, University Health Network manages health information using multiple information systems. The UHN Clinical Desktop is the primary information system by which health care providers can access comprehensive electronic chart clinical information. The UHN Clinical Desktop is integrated with many other on-line transactional processing systems, creating a landscape of event centric information. This architecture has many advantages from a systems management point of view, and is optimized for transactional applications. However, information remains isolated in disparate silos of data. One disadvantage is that analytical applications requiring data processing across information system involves a costly process. The Enterprise Data Warehouse (EDW) addresses this challenge by integrating information from the disparate data sources, in a central repository. The EDW is a persistent, qualified archive of discrete information collected from clinical and administrative systems used in the UHN patient care environment. The information stored in the EDW is organized primarily to support analytical applications. An overview of the EDW is provided in Appendix A.

1.2 Governance

The Enterprise Data Warehouse is an institutional resource and a corporate asset. Accordingly, the Enterprise Data Warehouse is governed by corporate policies related to confidentiality and disclosure of information.

2. Regulatory Environment

The Enterprise Data Warehouse was designed as an information service for multiple communities, including researchers and corporate decision makers. The Research Ethics Board (REB) and Corporate Privacy Office (CPO) oversee the policies and procedures governing access to information for research and administrative purposes, respectively.

2.1 Shared Responsibility

The protection of confidential information is an important responsibility that is shared among many information services and technology professionals at UHN. The management of complex, distributed information systems involves the coordination of procedures and practices by managers and administrators of networks, systems and databases.

The Corporate Privacy Office oversees the policies governing protection of confidential information and disclosure of data. The UHN Research Ethics Board has established a process for reviewing access to information for research purposes. Within this institutional framework, Research Informatics has worked with the Research Information Systems and Shared Information Management Services groups to establish safeguards to protect private and confidential information. These safeguards consist of standard operating procedures and technical measures by which Research Informatics manages the operation of the EDW.

2.2 Corporate Policies

As a corporate asset, access to the Enterprise Data Warehouse must comply with Corporate Policy 1.40.007, created to address UHN's compliance with privacy legislation. A complete description of corporate privacy information can be found on the CPO intranet site: http://intranet.uhn.ca/home/corporate_privacy_office/index.asp.

2.3 Access to Information Procedures

The process for research clients to obtain access to the EDW is outlined in Appendices F1 and F2.

2.4 Classification of Data

The Enterprise Data Warehouse contains many data elements, some of which are used to uniquely identify individuals. The Research Informatics group maintains the complete inventory of information stored in the EDW. Working with the Corporate Privacy Office, data elements containing private, confidential information were identified. Access to these patient-identifiable fields in the EDW is restricted:

i. Patient Name
ii. Medical Record Number (MRN)
iii. Mother's MRN
iv. Visit Number
v. OHIP Number
vi. Social Insurance Number (SIN)
vii. Address including postal code but excluding the forward sorting area (FSA)
viii. Telephone Number
ix. Employer

The business requirement to access to these fields needs to be explicitly justified for each access request. The Corporate Privacy Office reserves the right modify this list at any time.

3. Protection of Confidential Information

The standard operating procedures that have been established are designed to maintain the confidentiality, integrity and availability of information in the EDW. These procedures, encompassing physical, process, and technical controls enable appropriate access to the data contained within the EDW. Physical measures ensure that only authorized personnel carry out approved activities at the facility hosting the servers. Process level controls ensure that user requests are evaluated and approved under the current research and corporate approval stream. In addition, it includes robust practices for administering the EDW databases and applications that access it. Technical measures are implemented at the network, server, database and application levels. These controls pertain to technical configurations and enable an authorized user to access appropriate information at the authorized level of granularity. In a nutshell, the objective of all of the above controls is to allow the "right user" to access the "right information" at the "right time".

3.1 Physical Level Controls

3.1.1 Physical Environment

The EDW systems are housed on UHN property in a climate-controlled facility. Physical access to the facility is restricted to designated personnel in UHN Research Information Systems, and authorized individuals accompanied by UHN Security.

3.1.2 Physical Security

The EDW servers and storage devices are secured in the data room of the Centre for Global eHealth Innovation on the 4th floor of the R. F. Elliott Building in the Toronto General Hospital. During business hours, 9 am to 5 p.m., physical access to the EDW servers requires two levels security: 1) an electronic security badge, with appropriately configured access permissions, to enter the facilities where the data room is located; and, 2) a door key, which is restricted to designated members of UHN Research Information Systems and UHN Security. After business hours, access to the facilities is further restricted, requiring additional access to both the Centre for Global eHealth Innovation and R. F. Elliott Building using the electronic security badge system. UHN Security controls the electronic badges, and is responsible for the configuration of permissions to access these facilities. UHN security must receive written authorization from departmental managers before granting access to individuals.

3.1.3 Environmental Monitoring

The R. F. Elliott Building, the entrance and exit of the Centre for Global eHealth Innovation, and the facilities where the data room are located, are monitored by video surveillance. The temperature of the data room is continuously monitored. If the temperature of the room exceeds unsafe operating parameters, the UHN Environmental Services department is automatically notified to investigate.

3.2 Process Related Controls

3.2.1 Identity Management

Users need to be set up on the corporate or research networks prior to accessing the EDW. Once users are successfully connected to the corporate or research domains, they must then authenticate their identity before connecting to the EDW using a valid username and password. The Research Informatics group will only create accounts only after appropriate access documents are approved. Please refer to Appendix F1 and F2 for access procedures pertaining to the REB. Each user is given a unique username (generally their corporate T ID or research user name) that allows for pre-approved functions at the appropriate level of granularity.

3.2.2 Permissions

All research-based requests to access the EDW require approval from the UHN Research Ethics Board, as outlined in Appendix F. The approved access request dictates the user profile. The Research Informatics group creates the user profile and communicates an initial password to the user. The user is then advised to change their password.

Each user account is granted permissions according to their approved level of access. The user is only able to access information defined in their security profile, regardless of their method of connection to the EDW (SQL*Plus, Discoverer, ODBC, etc.).

Changes to access profiles are made based on new access requests only. Accounts with specific access periods are inactivated once their access period has ended or upon request from appropriate managers.

3.2.3 Operational Monitoring

3.2.3.1 Logging

Logging is performed at the administrative, database and application levels. At the administrative level, user requests are logged and information is recorded, including their user name, computer details, approved access details, report requests, copies of approved access requests and copies of emails in which user access is approved, are recorded.

At the database level, automatic logs of critical database events like start-ups, shutdowns, backups, system errors etc. is maintained. At the application level, logging is done by Oracle Discoverer, which documents the number of Discoverer queries each user has executed by business area and folder. This information is reviewed periodically to track usage and monitor database events.

3.2.3.2 Alerting

The Warehouse Administrator receives notification of critical events or system problems, such as:

- Database unavailable
- Disk space threshold limits exceeded
- Errors in refresh jobs etc.

The Warehouse administrator takes actions to correct the situation, and where necessary, the alert is escalated to one or more of the following groups:

- Director, Research Informatics
- Director, Research Information Systems
- User community

3.3 Network Level Controls

3.3.1 Network Management

The Research Information Systems (RIS) department of University Health Network operates and maintains the Research Network domain (uhnres.utoronto.ca). The server and storage infrastructure, which hosts the EDW, are connected to the Research Network. The networking equipment and connections between devices are physically secured in telecommunications closets, which are accessible only by designated RIS staff and personnel authorized by UHN Security. RIS actively manages the security of the network and has standard operating procedures in place to monitor the integrity of Research Network communications.

3.3.2 Authorized Devices

RIS uses industry-standard protocols to maintain the security of the Research Network. Network devices that are connected to the Research Network are assigned unique network names and addresses by RIS systems administrators. Users of these devices on the Research Network do not have administrative permissions to modify network settings.

3.3.3 Firewall Services

The perimeter of the Research Network is secured by a firewall, protecting the network from unauthorized connections. Devices on the Research Network are not visible to the outside. Authorized connections made to external devices are facilitated through network address translation.

3.3.4 Network Authentication

Users of the Research Network are assigned unique user names and are required to authenticate their identity using a confidential password. RIS has procedures in place to ensure that users employ strong passwords.

3.3.5 Intrusion Detection

Intrusion detection systems are part of the measures that RIS uses to monitor the Research Network and protect against unauthorized network communications.

3.3.6 Intrusion Protection

RIS has systems in place to ensure that the Research Network is protected from unauthorized devices establishing connections and initiating network communications. Devices on the network must be registered and assigned unique names and addresses by RIS, before they can connect to the Research Network.

3.3.7 Fault Tolerance

The Research Network is built on a redundant network architecture that allows communications to be dynamically re-routed around local failures in network segments.

3.4 Database Level Controls

3.4.1 Unauthorized Access

The user authentication mechanism employed by the database ensures that all connections to the EDW are established only after providing a valid username and password. To ensure password protection, passwords are sent over the network in an encrypted form. Other preventative security measures include password complexity and password change rules. Time sensitive user accounts expire after the end of their approved access period. In addition, user accounts can also be inactivated or deleted upon notification from appropriate channels.

3.4.2 Authentication

User authentication is performed by the database by comparing the submitted username and encrypted password to that stored in the database. Passwords are stored in an encrypted form in the database and cannot be viewed even by the Warehouse Administrator. The database connection is established only after successful authentication.

If three consecutive connection attempts are made using invalid username or password combination, the Oracle Discoverer or SQL Plus session shuts down and the unsuccessful connection attempts are logged in the audit trail for monitoring purposes.

3.4.3 Activity Log

The detection of inappropriate access is made possible by the EDW activity log. The audit mechanism in the database automatically records each event of user access, including along with the names of tables and fields that were accessed. An audit trail of user access

details (who, when and what) is recorded in the database. Reports on the audit trail are regularly reviewed to detect inappropriate access. Old records are archived on a regular basis.

3.4.4 Database Integrity

Data is validated for consistency between the source systems and the EDW on a periodic basis. All users are granted read-only access to EDW, which prevents data from being deleted or modified. The warehouse administrator ensures that all database changes are done in a specified and authorized manner. Different environments (development, integration, and production) serve specific functions, and allow for robust change management procedures. In addition, regular maintenance activities are scheduled to ensure database and system integrity.

3.4.5 Disaster Recovery

The data stored in the EDW is archived on a periodic basis by Research Information Systems, capturing the state of the EDW at a given point in time. The media on which the EDW is archived, are stored outside the facility in which the EDW storage resides. These archives can be used to recover the database in the event of loss of mirrored disk pairs, data corruption or catastrophic failure. The database operates in archive-mode, which ensures point-in-time recovery.

3.4.6 Availability

EDW is hosted on a robust and scalable hardware and software platform. The server environment has excess capacity to meet near future needs. The operating environment ensures that EDW has high availability for our user community.

3.4.7 Reliability

The server environment hosting the EDW is configured for device (hard disk and sever) redundancy. Disks are mirrored using RAID1 configuration, so that failure of any one of the mirrored disks does not impact the availability of the warehouse. In addition, the operating system uses virtual server environments, which allows for failover to another virtual server. The EDW environment configuration is illustrated in Appendix E.

3.4.8 Database Performance

Automated monitoring processes are in place for ensuring system availability, storage utilization and vital database performance statistics.

3.5 Application Level Controls

Business Areas (BA) in the EDW contain information and reporting workbooks that have been pre-configured based on user needs. These business areas are accessed using the Oracle Discoverer application. The Oracle Discoverer Administrator is responsible for application-level control over the information accessible by any given Discoverer user. Upon receiving the appropriate approval, the Administrator creates an individual account and grants the user access to specific, approved EDW BAs.

3.5.1 Implementation of Security Measures

Based on the classification of data elements, database fields containing confidential information were identified. In order to allow access to the EDW, without exposing confidential information, a view of the database was constructed which does not contain any

of the patient-identifiable fields. This view of the EDW contains de-identified records. A second view, containing the 9 patient-identifiable data elements was also constructed. The result is that there are two possible views of the EDW:

i. Public End User Layer, which contains only de-identified data;
ii. Private End User Layer, which contains patient-identifiable data;

In the course of administrating access permissions, the Warehouse Administrator will assign access to BAs using the appropriate view of the EDW.

3.5.2 Administration of Access

Research Informatics is responsible for the operation of the EDW and the administration of user accounts. Authorization of access to information for research purposes is the responsibility of the Research Ethics Board.

3.5.3 Authorization

Authorization of investigator requests will come from the UHN Research Ethics Board, which has established a procedure for reviewing requests for information for research purposes. The procedure and the accompanying form are found in Appendices F1 and F2, respectively. Investigators interested in gaining access to data for study should consult the Research Ethics Board website on the Research Intranet (http://intranet.uhnres.utoronto.ca).

3.5.4 Account Creation

The Warehouse Administrator is responsible for creating user accounts. A unique account will be created for each investigator associated with REB approved access. Each user account is created based on the corporate T ID or research user name of the user.

3.5.5 Access Permissions

After the account has been created, the Warehouse Administrator will grant access to either the Public or Private End User Layer, as is appropriate given the investigator's approved request. Once the appropriate level of confidentiality is granted by the Warehouse Administrator, the Discoverer Administrator will then assign access to appropriate business area(s). At this point, the client can access the business area using Oracle Discoverer.

3.5.6 Account Expiration

When the scheduled access period expires or a client leaves their role, the Warehouse Administrator will de-activate the user's account. Authorization of account de-activations for any other purpose will come from the EDW Manager.