
National Patient Information Reporting System: National Data Warehouse

NDW General Data Mart Getting Started Guide

Current Version 8.0

Created: 2/15/2012

Published: 6/4/2012

Author: jvosburg



Department of Health and
Human Services

Indian Health Service

Office of Information
Technology (OIT)

Contents

Contents	ii
Version Control	iii
Introduction.....	1
About this Guide.....	1
Accessing the General Data Mart.....	2
GDM Initial Access Instructions	3
Reset Your Password	3
Download the DB2 Client.....	4
Connection Configuration.....	5
Contact Information	5
Access Privileges.....	6
Database Access.....	6
Area Access	7
General Data Mart Organization	7
Patient Registration Data	8
Non-PII Registration Tables	9
Encounter Data.....	9
Reference Data.....	11
Administrative Data	11
Data Retrieval – Best Practices.....	11
Views	11
Appendix A – Table Structure.....	12
SECURITY_TAG Column.....	12
Data Model.....	12

Version Control

Version	Date	Notes
1.0	September 2005	Initial Draft
2.0	October 2005	Initial Release
2.1	November 2005	Changes based on customer feedback.
2.2	December 2007	Sub mart of Hollywood, add REF info
2.3	February 2008	Annual Review, removed Appendices (A-E) as information is available at the NDW web site at the following link: http://www.ihs.gov/CIO/scb/metadata/ FY08 Bridge Contract deliverable (1.7.6); COTR accepted April 11, 2008
2.4	August 2008	Update for export_info and integrity ID additions by area access
3.0 Draft	March 2009	Add/Remove: <ul style="list-style-type: none"> • schemas for the Restricted Personally Identifiable Information (PII) Data Requirements • data element that allows control of Area security • schemas for each version of past userpop reports run from NDW data • remove ssn from Pat_reg, userpop and enctrss tables • add new tables; pat_reg_ssn, userpop_ssn and enctrss_ssn COTR approval April 3, 2009
3.1	November 2010	Annual Update
4.0	June 2011	Final
5.0	February 2012	Updated metadata with missing Target Folder and Target Path information.

Version	Date	Notes
6.0	February 2012	Added entirely new section on resetting AIX password and installing the DB2 Client.
7.0	March 2012	Changes finalized and approved.
7.2	March 2012	Additional update: revision to a paragraph related to DUP_FG.
8.0	June 2012	Changes finalized and approved.

Introduction

The National Data Warehouse (NDW) General Data Mart (GDM) contains a sub-set of the NDW registration and encounter data. The contents of this data mart include:

- All current registration data contained in the data warehouse
- Full registration data which **includes** Personally Identifiable Information (PII)
- Non-PII registration data, which **excludes** Personally Identifiable Information (PII)
- Current user population table.
- All current encounters contained in the data warehouse
- All reference tables
- Meta data, which contains information about the tables and columns included in the database
- Administrative data, including a timestamp for Mart refreshes and also information pertaining to file exports

The NDW General Data Mart is refreshed per the *General Data Mart Service Level Agreement*. The General Data Mart also includes the Legacy Data Mart. Data in tables associated with the Legacy Data Mart is not refreshed.

About this Guide

The *NDW General Data Mart Getting Started Guide* was developed to give experienced users a fundamental understanding of the General Data Mart. It is written specifically to address the needs of those users with more advanced statistical, analytical and programming skills.

This guide does not provide complete documentation of the data mart, database use or concepts, nor is it a programming language tutorial. It does, however, include detailed instructions for initial setup of a client machine to enable GDM access, including how to change the AIX password and how to install the DB2 client.

Accessing the General Data Mart

System access is enterprise compliant, allowing various environments to access the database, including ODBC, JDBC, OLE, and CLI. Software applications that can access databases using one of these standard protocols can connect to the General Data Mart. NPIRS will assist in obtaining a connection to the GDM but is not responsible for access programs.

The steps for setting up a machine for first-time access to the GDM are provided in the following section.

GDM Initial Access Instructions

Reset Your Password

1. Open your web browser and connect to `http://export /ssh`.
2. The Security Warning message (shown below) will appear.



Figure 1 – Screen capture of the “Security Warning” message.

3. Allow the Java application to run by clicking on the **Run** button.
4. When the Connection Profile window appears (shown below), enter the server’s IP address in the “Hostname” text box, and type your username in the “Username” text box.

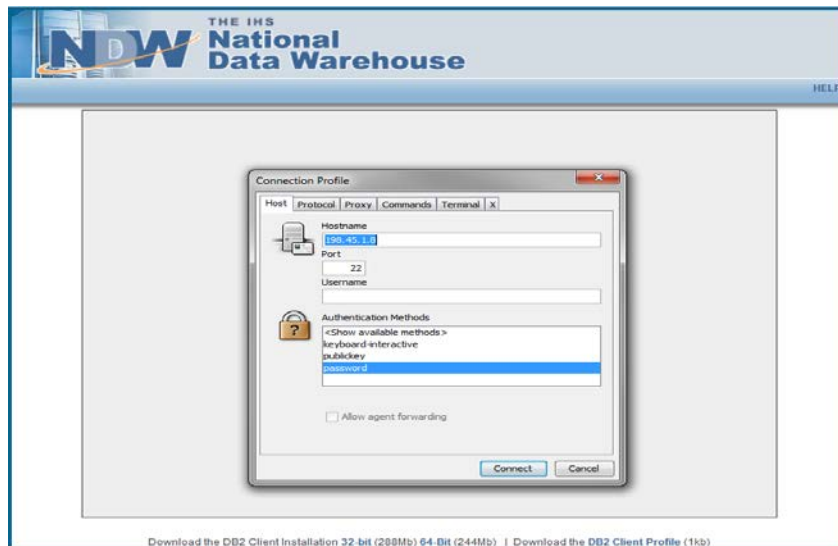


Figure 2 – Screen capture of the “Connection Profile” window.

5. Select “password” for Authentication Method; then click the **Connect** button.
6. A message will appear prompting you to accept the host key fingerprint. Click on the **Yes** button.

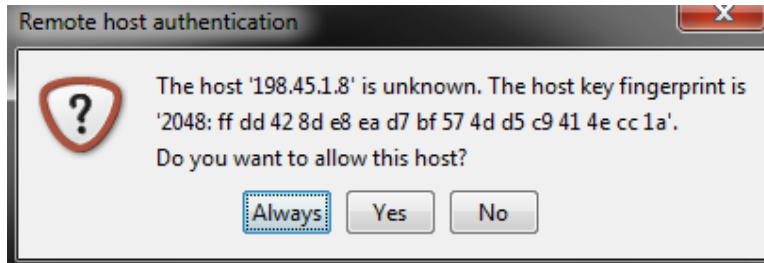


Figure 3 – Screen capture of the “Remote Host Authentication” window.

Download the DB2 Client

Links to the Client are on the bottom of the NDW SSH Page. Simply click on the link to download either the 32 bit client or the 64 bit client.

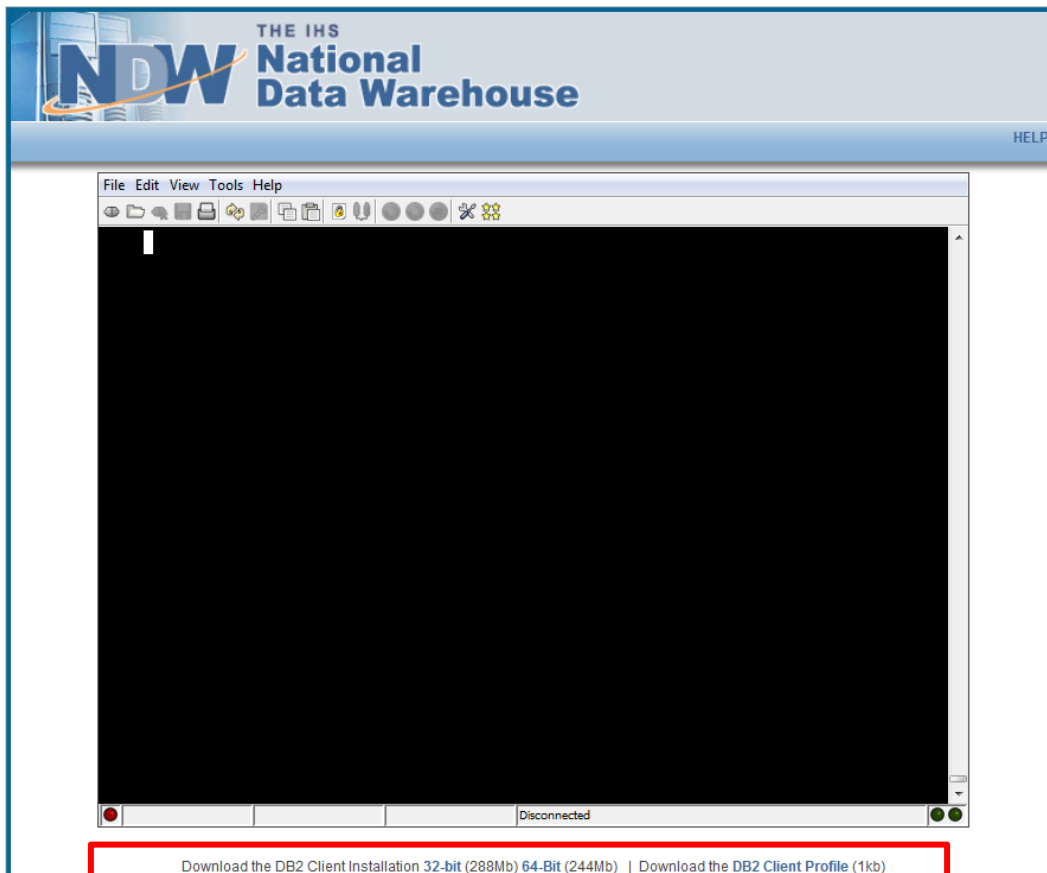


Figure 4 – Screen capture of the “Remote Host Authentication” window.

Connection Configuration

Protocol: TCP/IP
IP Address: 198.45.1.8
Host Name: bilbo.hqw.ihs.gov
Database Name: HOLLYWD
Remote Instance Name: db2hol

For users requiring access to a database other than HOLLYWD, contact the NDW staff for appropriate connection configuration parameters .

Contact Information

To request access to the IHS General Data Mart,

Fill out an appropriate ITAC form or if you need assistance, send an email to: support@ihs.gov..

An invitation to the next General Data Mart Users Group meeting can also be requested by sending an email to support@ihs.gov. This group meets on a regular basis.

Access Privileges

Database Access

Access to the General Data Mart is restricted to authorized users. Authorized users currently include IHS departments and programs (e.g. EPI, IHPES, and Diabetes), some external research-driven corporate entities, and some Area Statistical Officers.

Authorized users who are granted access to the General Data Mart will be assigned one of the following security access levels:

National Level 1 Allows a user to view all data for all Areas.

National Level 2 Allows a user to view partially cleansed data for all Areas.

Area Level 1 Allows a user to view all data within their specified Area.

Area Level 2 Allows a user to view partially cleansed data within their specified Area.

Below is a graphical representation of the four different levels of access.

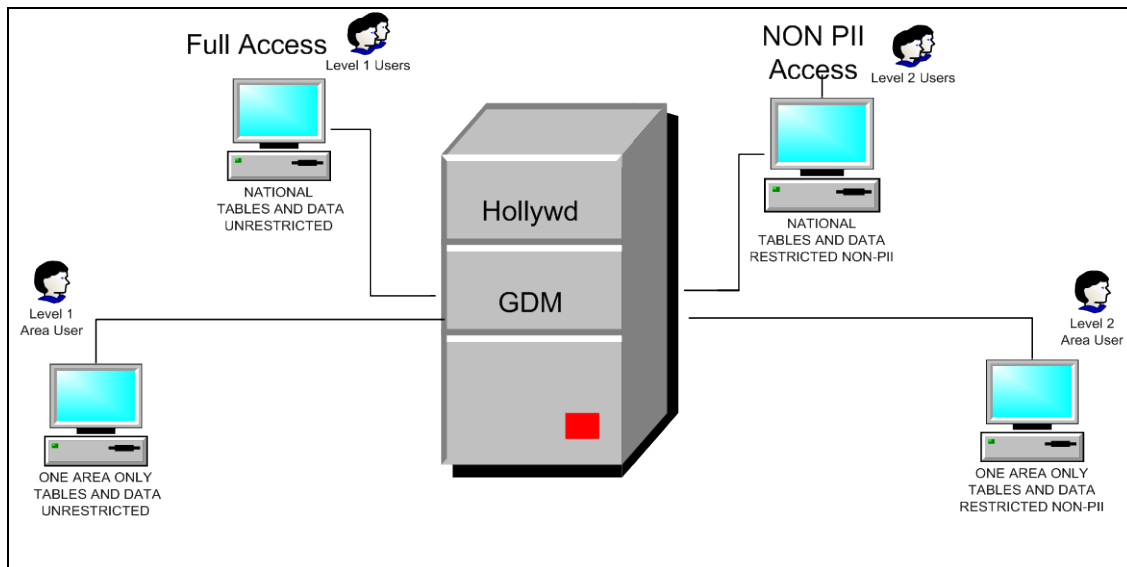


Figure 5 - Illustration of the different levels of national and area access

In all cases, authorized users have **Read-Only** privileges. Additional security levels may be assigned or created in the future.

The term “partially cleansed” refers to data that has had the PII removed. Tables containing partially-cleansed data may be easily identified by the names of their schemas, which will contain the suffix “_NP” (for “non-PII”). For example, the REG_NP schema contains Registration data that is *not* PII sensitive (i.e., has been partially cleansed). The ENCTR schema generally does not contain PII data, whereas the ENCTR_NP and ENCTR_PII schemas do.

The fields and layout of the General Data Mart tables are located in the document titled “*GDM Schemas and Tables/Views/Nicknames*” which is accessible from the website at: http://www.ihs.gov/NDW/index.cfm?module=dsp_dqw_mq4.

Area Access

Area users may access only Area based data from the ENCTR, REG or REG_NP schemas and data relevant to their security access. Area Access is controlled by the Security Administrator and based on the user’s ID.

General Data Mart Organization

The General Data Mart is organized into the following “groups” that separate different kinds of data into their own spaces.

- REG contains the registration tables created to store information obtained from patient registration records received by the National Data Warehouse.
- REG_NP is a subset of the registration data excluding PII data. Tables in REG_NP are used as the base tables for the views/MQs for the various security views.
- ENCTR contains the encounter tables created to store information about patient encounter records received by the National Data Warehouse.
- REF contains both the Standard Code Book tables and the IHS code sets used by the National Data Warehouse (NDW). IHS Standard Code Book tables are available on our web site: <http://www.ihs.gov/CIO/scb/>. For a complete list of the Reference tables, see the document “*GDM Schemas and Tables/Views/Nicknames*”.
- META tables contain basic information on the structure and definition of columns and tables used in the mart. This information may be accessed via the meta-data web site: <http://www.ihs.gov/CIO/scb/metadata/>.
- SYSIBM (SYSCOLUMNS, SYSTABLES) also contains meta-data about the mart.
- ADMIN contains two tables:
 - LAST_ETL – contains the dates of the extracts to refresh the mart.

EXPORT_INFO_DATA – contains information about file exports.

LEGACY tables contain static data from the legacy NPIRS system. This consists of data collected prior to the startup of the NDW database. For additional information concerning the legacy tables and their usage, see the “*Legacy Data Mart Getting Started Guide*”. This is historical data (Pre-NDW) and is not refreshed. Standard NPIRS supported tables and columns in the GDM can be viewed in the document “*GDM Schemas and Tables/Views/Nicknames*” at http://www.ihs.gov/NDW/documents/NDW_DataAU/GDM_Schemas_TablesViewsNicknames_V2.0.pdf.

Additional information on the GDM, including minutes from the latest GDM User’s Group Meeting, is located on the General Data Mart home page at <http://home.ihs.gov/NDW/GDM/index.cfm>.

Patient Registration Data

The **Patient Registration** tables contain information about the *individual* patients registered in the various IHS and tribal facilities.

Note: A patient may be registered at more than one facility, or more than once at the same facility, and therefore may have multiple registration records. No un-duplication is applied. The REG.USERPOP AND REG_NP.USERPOP tables both contain a data element called INTEGRITY_ID. This column contains a unique identifier assigned to all registration records included in a given version of USERPOP data and considered the same person, using approved business rules.

Within the REG schema, the PAT_REG and DEMOGR tables contain basic registration information about the patient. Typically, PAT_REG holds more static patient information, such as Birth Date, Gender, and SSN. DEMOGR holds more changeable patient information, such as Name, Address, and Community of Residence.

The CHART, ALIAS, and INSUR_ELIG tables contain additional information about the patient, if applicable, and there may be multiple records for that patient. Records in these tables utilize a sequence number field (SEQ_NBR). The sequence number represents the order in which the data was received by the NDW.

The primary key used by the Patient Registration tables is **REG_ID**. A Registration Identifier (REG_ID) is the primary data element within all Registration tables that uniquely identifies each registration record in Registration tables within REG and REG_NP schemas. This key is the unique identifier for the Patient Registration record, and is used to link all of the patient's registration-related data. The REG_ID is generated by the NDW to ensure uniqueness of the registration within the database. The REG_ID field is found in each of the Registration tables and may be used to retrieve registration information for a single registrant from multiple tables.

Note: Some encounter records are received without registration information. These records will have REG_ID = 0 and cannot be linked to registration records.

Encounter information (discussed in the next section) can be linked to individual patient registration records for reporting purposes via the REG_ID (when a registration record exists). The REG_ID is the typical method of linking encounters and registration data, but other methods may be used, depending on the intent of the query.

Non-PII Registration Tables

The REG_NP schema has many of the same tables as the REG schema; however, any data that is considered to be PII (information that can directly identify a patient) has been excluded from the Non-PII schema in order to uphold security requirements for the General Data Mart users. Some PII data, such as birth dates and death dates, have been retained in order to assist the user in preparing reports.

Details related to the REG_NP schema can be viewed in the "Registration Non-PII Tables" section of the document "*GDM Schemas and Tables/Views*".

Encounter Data

The ENCTRSS table contains the basic information about encounters. Other tables in the ENCTR schema contain more specialized information, such as dental and contract data. For example, the ADA_PROC table contains standardized American Dental Association codes related to a dental visit. Explanations of these tables and what they contain can be found on the IHS NDW Meta Data web site at: <http://www.ihs.gov/CIO/scb/metadata/>

The data in encounter tables represent the most current version of the record sent in by each source. Some CHS (Contract Health Service) encounter related data may be received from multiple sources and may portray a different level of granularity (visit record; purchase order record; claim record).

Note: The WL_DUP table contains encounter ids and flags indicating whether a received record is a duplicate using business rules established by DPS for counting workload. The DUP_FG may be utilized for certain reporting purposes, but in general should not be used when searching for certain clinical conditions. For example, using the DUP_FG when searching for certain diagnosis codes may result in not retrieving all records meeting the search criteria.

For a complete listing and description of the Encounter tables, as well as models with physical and logical names, go to the IHS NDW Meta Data web site:

<http://www.ihs.gov/CIO/scb/metadata/>.

The NDW Encounter Identifier (ENCTRSS_ID) is a key data element that distinctly identifies each encounter record in Encounter tables within the ENCTR schema. The ENCTRSS_ID is generated by the NDW to ensure uniqueness of the encounter within the database. The ENCTRSS_ID field is found in each of the Encounter tables and may be used to retrieve encounter information stored in multiple tables.

A sequence number (SEQ_NBR) is stored for multiple child records of the same encounter. This number is the order in which NDW receives the data.

The main Encounter table (ENCTRSS) contains REG_ID field which indicates the NDW registration identifier for the encounter data.

Encounter records are usually linked to a patient registration in the Registration schema (REG) by registration identifiers (REG_ID) stored in the main encounter table (ENCTRSS). Some encounter records will not have registration IDs due to timing issues in receiving encounter and registration data by NDW. The registration ids (REG_ID) for those records will default to 0. Prior to each GDM refresh there will be an attempt to link these unlinked encounters to a newly received registration.

Reference Data

The reference tables contain descriptions and other pertinent data concerning the codes utilized in the NDW. The reference tables are in the REF schema. The code values are generally listed as a column name ending with _CD. For example, the FACILITY table contains the facility codes (FAC_CD) and names (FAC_NM) for facilities. In addition, some tables contain a flag indicating whether data in a column is workload reportable.

The reference tables reflect codes that are used in the RPMS and NDW system. They may not be the most current list since the NDW does not receive regular updates from RPMS for most tables. Check with your local system administrator for a complete listing.

For a complete list of the tables with descriptions, see the document entitled “*NDW Reference Tables Reference Guide*”.

Administrative Data

Administrative tables are referenced by the ADMIN schema. The LAST_ETL table contains the dates of the extracts that were performed to refresh the mart and is used for informational purposes only. The EXPORT_INFO_DATA table contains data about the file exports, including EXPORT_ID and various dates.

Data Retrieval – Best Practices

Views

Database views may be used to organize data in alternative ways, depending on user requirements. Keep in mind that there are two levels of view security – “private” and “public” - which enable the system to distinguish between users who are only authorized to read the data contained in a view, and those who are authorized to alter the contents of the views themselves.

- 1) **Private** – If the view is created using the creator’s user-id as the schema (this is the default), the view and data will be made accessible only to the person creating the view.
- 2) **Public** – If the view is created using ENCTR, REG, REG_NP, or NATL, the view and data will be accessible to users having the appropriate access.

Views should be used only when they are expected to remain persistent (i.e., used over and over, to resolve very complex SQL code), or if they are expected to be used in a stored procedure. As a rule, a user is expected to drop (delete) views that are no longer needed.

Appendix A – Table Structure

The tables below are used to fulfill the security requirements for the General Data Mart (GDM). A schema for Registration tables with Non-PII (REG_NP) houses the tables that have been partially cleansed of Personally Identifiable Information (PII). The REG_NP schema is comprised of registration tables for use by both National and Area Level 2 users.

The following tables are directly related to the implementation of security on the GDM:

- REG_NP.PAT_REG
- REG_NP.INSUR_ELIG
- REG_NP.USERPOP
- ENCTR.ENCTRSS_SSN*
- REG_NP.DEMOGR
- REG_NP.CHART
- REG.PAT_REG_SSN*
- REG.USERPOP_SSN*

** The REG.PAT_REG_SSN, REG.USERPOP_SSN and the ENCTR.ENCTRSS_SSN tables allow only limited user access, based on the individual user's security access level.*

SECURITY_TAG Column

All tables in the REG schema contain a binary security tag column (SECURITY_TAG). This column cannot be used for reporting; its purpose is to enable the Security Administrator to apply the appropriate security access levels to the tables.

Data Model

All the tables and columns in the GDM can be viewed in the document entitled “*GDM Schemas and Tables/Views/Nicknames*” located at http://www.ihs.gov/NDW/documents/NDW_DataAU/GDM_Schemas_TablesViewsNicknames.pdf.