
National Patient Information Reporting System: National Data Warehouse

NDW General Data Mart

Getting Started Guide

Version 2.3

February 2008



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
Connection Configuration.....	2
Contact Information	2
Database Access.....	3
Access Privileges	3
General Data Mart Organization	4
Patient Registration Data	5
Encounter Data.....	6
Reference Data.....	7
Administrative Data.....	7
Data Retrieval - Best Practices	7
Views	7
SAS Libraries.....	8

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. FY08 Bridge Contract deliverable (1.7.6); COTR accepted April 11, 2008

Introduction

The National Data Warehouse (NDW) **General Data Mart** 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
- All current encounters contained in the data warehouse
- All historical 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, which is information concerning when refreshes were performed.

The NDW General Data Mart is refreshed per the *HOLLYWD Database and Associated Data Marts Service Level Agreement, Version 2.0*. The General Data Mart also includes a copy of the Legacy database tables. These tables are 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 and analytical skills.

This guide does not provide complete documentation of the data mart, database use or concepts; nor is it a programming language tutorial. Examples provided within this manual use standard DB2 SQL syntax and may not work for you, depending on your database query application, but they should provide a starting point to understand better the database structures and the data contained therein.

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, including SAS.

Connection Configuration

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

Contact Information

To request access to the IHS General Data Mart, call or email

Stanley P. Griffith, MD, FAAFP
Manager, IHS National Data Warehouse Project
Indian Health Service, Office of Information Technology
Phone: 1-505-248-4144
E-mail: stanley.griffith@ihs.gov

Or send an email to: ITSCDataWarehouse@ihs.gov

For information about workstation SAS, call or email

Diana Roberts, Biostatistician
Indian Health Service, Office of Information Technology
Phone: 1-505-248-4896
E-mail: diana.roberts@ihs.gov

For information about server SAS, call or email

Larry Layne, Epidemiologist
Indian Health Service, Office of Information Technology
Phone: 505-248-4132
E-mail: larry.layne@ihs.gov

To promote a collaborative environment between the NDW staff and the users of the General Data Mart, the General Data Mart Users Group meets on a regular basis.

Database Access

Only *authorized* users are allowed access to the General Data Mart, and will be assigned one of the following security levels:

- **National Level** access allows a user to view all data for all Areas.
- **Area Level** access allows a user to view all data within their specified Area.

In all cases, authorized users have **Read-Only** privileges.

Access Privileges

Each table is presented in a “view,” based on the user’s designated security access. Area access privileges are determined by a 3-character code representing the Area that the user is authorized to view. National access privileges are determined by the NATL code, which allows access to all data for all Areas. The following table lists the access codes:

Code	Description
AKA	Alaska
ABR	Aberdeen
ALB	Albuquerque
BJI	Bemidji
BIL	Billings
CAO	California
NSA	Nashville
NAV	Navajo
OKC	Oklahoma City

Code	Description
PHX	Phoenix
POR	Portland
TUC	Tucson
NATL	All Area access

For example, a user with Area only access to the Tucson Area data would use **TUC** as the prefix to the PAT_REG table to query the Tucson Patient Registration table:

“**TUC.PAT_REG**”

A user allowed access to all data regardless of Area would use **NATL** as the prefix to the PAT_REG table to query the Patient Registration table:

“**NATL.PAT_REG**”

Further, a NATL access user may also query a certain Area’s data by using the 3-character Area code, for example, OKC, as the prefix to the Patient Registration table to view only Oklahoma’s patient registration data.

Note: If a user with Area privileges (e.g., OKC) were to attempt to use NATL or an Area prefix other than their own, access to National data and the other Area data would be denied.

General Data Mart Organization

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

- REF contains the code sets (reference tables) used by the National Data Warehouse (NDW). The primary reference tables are available at the IHS Standard Code Book web site: <http://www.ihs.gov/CIO/scb/>

For a complete list of the Reference tables, see the *NDW Reference Tables* document.

- META contains the metadata about the mart, such as data dictionary information. Data models, the tables list and data dictionary may be accessed via the meta data web site: <http://www.ihs.gov/CIO/scb/metadata/>.
- SYSIBM (SYSCOLUMNS, SYSTABLES) also contains metadata about the mart.

- ADMIN contains the dates of the extracts to refresh the mart (LAST_ETL)
- XXX, where XXX equals the 3-character code of the Area you are authorized to view; or NATL, if you have access to all data for all Areas. For a complete list of the Area codes, see the “Access Privileges” section on page 3.
- LEGACY tables are static copies of the data from the legacy NPIRS system, containing data 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* Version 1.0.

Patient Registration Data

The **Patient Registration** tables contain key 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 patient registration records.
No un-duplication is applied.

Patient Registration data is stored in multiple tables. For a listing and description of these 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 PAT_REG and DEMOGR tables (both of which are required for adding a registration record to the NDW) are the two tables that contain “base” information about the patient. Typically, PAT_REG holds more static patient information, such as Birth data, Gender, and SSN. DEMOGR holds changeable patient information, such as Name, Address, and Community of Residence.

The primary key used by the Patient Registration tables is **REG_ID**. 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 **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**) to order multiple records with a given REG_ID.

Encounter information (discussed in the next section) is linked to individual patient registration records for system purposes via the REG_ID field. The REG_ID field is found in each table of the Registration schema and may be used in table join operations. 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.

Encounter Data

The ENCTRSS table contains the basic information about encounters. All other encounter-related tables contain optional detail information that not all encounters will contain. Other encounter tables contain more specialized information, such as dental and contract data. The table name is a good indicator of the type of data stored therein.

Note: The NDW receives information about a visit from multiple sources (e.g., PCC, HCHS/MIS, CHS FI, etc.), and therefore may hold more than one row per encounter. **No un-duplication is applied.**

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 primary key used by the Encounter tables is ENCTRSS_ID. This key is the unique identifier for the Encounter record, and is used to link all of the data for that encounter. The ENCTRSS_ID is generated by the NDW to ensure uniqueness of the encounter within the database.

Multiple records for a given table use the SEQ_NBR field to order the records by ENCTRSS_ID.

Encounter records are always linked to a patient registration in the PAT_REG table. For example, ENCTRSS.REG_ID points directly to a record in PAT_REG, which indicates who the patient was for that encounter.

The CURR_ENCTR_SS_FG column set to “Y” identifies the most current encounters. Since the NDW holds a history of all encounters exported by the site, this flag indicates which encounter is the most recent, or current, received.

Current encounter data is stored in the Encounter (ENCTR) tables.

All non-current and deleted encounter data is stored in the Encounter History (ENCTR_HIST) tables. The ENCTR_HIST tables are provided as supplements only; they are utilized administratively to ensure synchronization between the Hollywd database and the NDW.

Reference Data

The reference tables, also known as lookup 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. Additional fields in each table are dependent on the type of code that is referenced.

For example, the FACILITY table contains the ASUFAC (FAC_CD) and name (FAC_NM) for facilities. In addition, some tables contain information as to whether the particular code is workload reportable.

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

Administrative Data

The LAST_ETL table contains the dates of the extracts that were performed to refresh the mart and is used for informational purposes only.

Data Retrieval - Best Practices

Views

Database views may be used to organize the data in alternative ways, depending on user needs. When doing so, keep in mind that there are two different kinds of views, and this affects the security for the view (who is authorized to view the data contained, and who may alter the views).

- 1) **Private** - If the view is created using the creator's userid as the schema (this is the default), the view and data will be accessible only to the person creating the view.

- 2) **Public** - If the view is created using ENCTR, REG, NATL (or one of the area prefixes), the view and data will be accessible to others having National (or area) access, and may become available to Area level users in the future following security review.

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 expected to be used in a stored procedure. In general, a user is expected to drop (delete) views that are no longer needed.

SAS Libraries

Temporary datasets may be stored for a short period of time in the usr/tmp directory on the server until they are moved to the workstation. For more information, see the *HOLLYWD Database and Associated Data Marts Service Level Agreement, Version 2.0*.