
National Patient Information Reporting System: National Data Warehouse

NDW Workload/User Population Data Mart Technical Guide

Version 1.2

February 2008



Department of Health and
Human Services

Indian Health Service

Office of Information
Technology (OIT)

Contents

Version Control	iii
Requirements	1
Overview of the User Population/Workload Process	2
Design Parameters	4
Data Parameters	6
System Environment	6
User Population/Workload (UP/WKLD) Data Mart	6
NDW Production Data Mart	6
IHS Intranet – Web Reporting	7
Security	7
Data Transfer Process	7
User Population.....	8
Workload.....	10
Reporting.....	11
Archival	12
Appendix A: Data Transfer Process Diagrams	13
User Population Process	13
Workload Process	14

Version Control

Version	Date	Notes
1.0	August 2007	FY07 Contract Deliverable (1.12.6) Approved August 24, 2007
1.1	December 2007	Added business logic modification to implement two RPMS PCC edits (H8, A6), which will prevent encounters with no POV (purpose of visit) diagnosis code from being workload reportable (Page 11). FY08 Bridge Contract Deliverable (1.7.6) Approved January 3, 2008
1.2	February 2008	Annual review and update Removed Appendix A, Schemas & tables into separate docs, as referenced on page 6. FY08 Bridge Contract Deliverable (3.1.3) COTR approved April 10, 2008

Requirements

The government has requested that the National Data Warehouse (NDW) provide a Data Mart dedicated to producing User Population and Workload reports. This Data Mart will be comprised of data from the NDW production database that will satisfy the requirements for the production of workload and user population reports, their verification, and distribution.

The User Population/Workload Data Mart will be the sole source of data for user population (userpop) and workload reporting, utilizing user population and workload data that has been extracted from the NDW production database.

Special reports may be run at customer request, which may require different data subsets.

IMPORTANT

The Userpop/Workload Data Mart database contains a copy of **LIVE** production data.

No scrambling, cleansing, encryption; or other methodology is used to disguise patient identifiable data.

Overview of the User Population/Workload Process

The following diagram provides a general flow of the data from the NDW production database to the User Population/Workload (UP/WKLD) Data Mart and the reporting process.

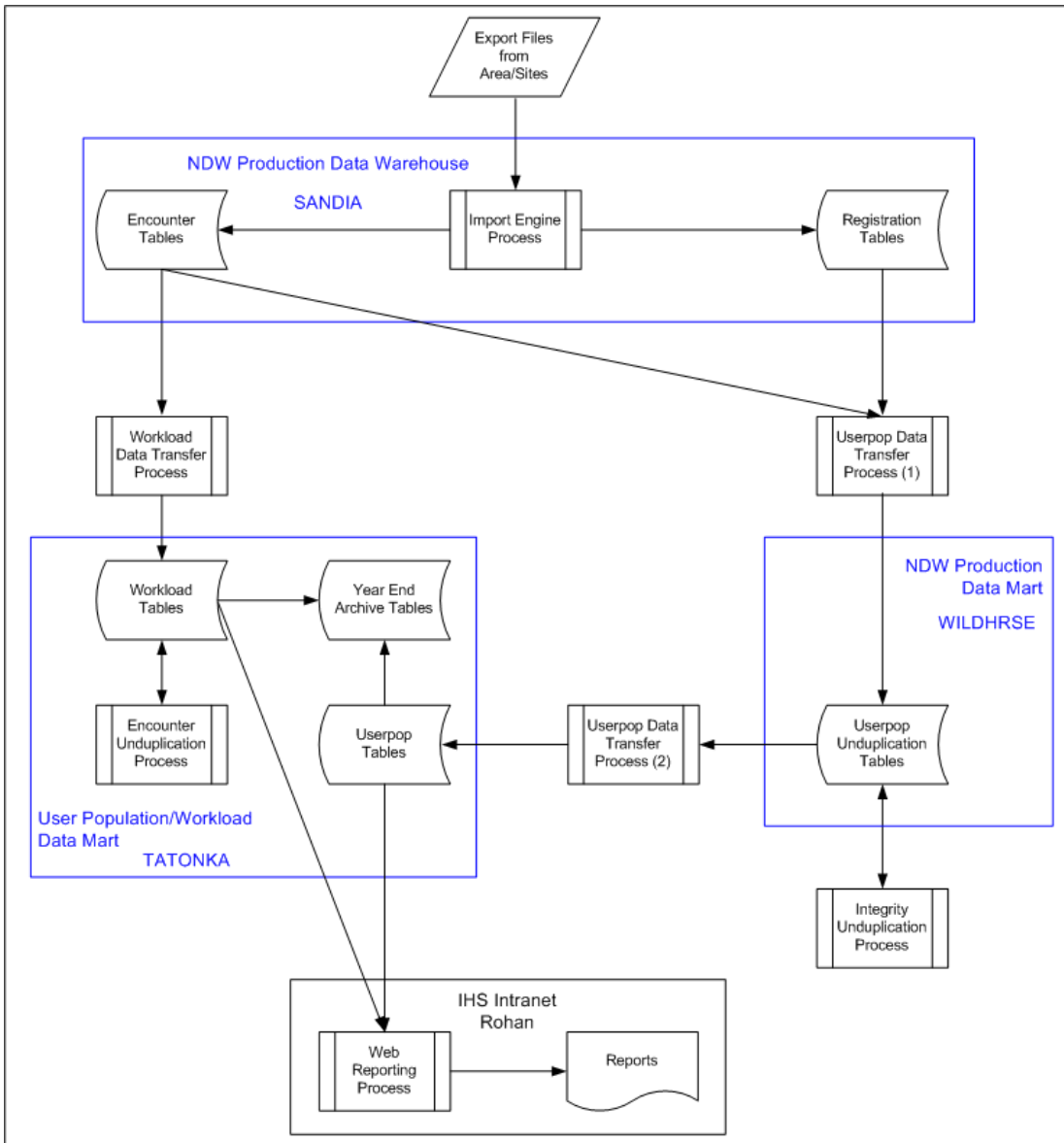


Figure 1. User Population/Workload Process

As illustrated in Figure 1,

- **NDW Production Data Warehouse (SANDIA)**

This database contains all registration and encounter data sent to the NDW via export files from the Areas and sites.

- **Registration Tables** include Patient data (e.g. name, demographic, chart, insurance).
- **Encounter Tables** include Encounter data (e.g. location, provider, procedures, etc.).
- **Import Engine Process** is the application that loads the export file data into the warehouse.

Integrity Software is the application that provides patient unduplication results for user population extract runs. AIX Version 3.10 is currently used and new versions are referred to as Quality Stage.

- **Workload Data Transfer Process**

This process selects encounter data that are applicable for workload reporting and transfers the data to the Workload/User Population Data Mart tables.

- **Userpop Data Transfer Process (1)**

This process selects registration data that are applicable for user population reporting and transfers the data to the NDW Production Data Mart (WILDHRSE) for unduplication.

- **NDW Production Data Mart (WILDHRSE)**

This database was created to allow the unduplication process and other computations to be performed against production data without interrupting the regular production transaction processing. The NDWMART schema within this database houses unduplicated production data that must be made available to other data marts.

The **Userpop Unduplication Tables** in the NDWMART schema are loaded with valid registration data. That data is then processed through the Integrity application, which determines duplicates. The unduplication tables are then updated with survivorship information.

Unduplicated registration data will be available to other marts per the government requirement to maintain and make available a retrievable record of the enterprise and area-wide Integrity unduplication of registration patients for:

- each approved annual User Population Report

- each reporting period per Government Performance reporting Act (GPRA) regulations
- each draft User Population report
- **Userpop Data Transfer Process (2)**

The User Population data is transferred to the User Population/ Workload Data Mart (TATONKA) for reporting.
- **User Population/Workload Data Mart (TATONKA)**

This data mart contains all data needed for workload and user population reporting.

 - **Workload Tables** include all encounter data needed for workload reporting
 - **Userpop Tables** include all registration data needed for User Population reporting. The data in these tables are versioned, so that previous User Population snapshots are available.
 - **Encounter Unduplication Process** determines encounter duplicates and updates the workload tables with survivorship information.
- **Year End Archive Tables**

Once the final reports have been produced for a fiscal year, the summary information for that year's reporting data is archived to the WILDHRSE database. Additional information can be found in the *UserPop/Workload Year-End Archive Process* document.
- **Web Reporting Process**

All official reports are available at the IHS National Data Warehouse intranet web site for viewing by authorized users. These reports meet government requirements to provide reporting as contracted, and other reporting as requested.

Design Parameters

The User Population/Workload (UP/WKLD) Data Mart and processes were designed according to the NPIRS Business Rules. These rules, as they apply to the legacy system, are available in the *NPIRS Basic Business Rules*, Version 1.4 (May 2004). An updated document will be produced to reflect the new data names and processes implemented for the NDW.

The basic business logic, as stated, was applied at the inception of the NDW. Some workload logic has been modified per the government's request since then; see specifications for these two changes in the following documents:

- *Dental Encounter & Workload Specification V1.2*
- *Workload Flag POV Specification V1.0*

User Population:

- The Registration record unduplication process takes place within the NDW Production Data Mart (WILDHRSE), using the Integrity software. This process takes place for each scheduled user population reporting cycle, and the resulting data is then transferred to the User Population/Workload (UP/WL) Data Mart in Tatonka for reporting purposes.
- User Population reports are produced per the approved schedule, which can be found in the *Schedule for Production of FY xxxx User Population Estimates*, published by the Office of Public Health Support, Division of Program Statistics, Statistician and Director.

Workload:

- Encounter unduplication is performed after the load to the User Population/Workload Data Mart.
- Workload reports are produced per the approved schedule, which can be found in the *Schedule for Production of FY xxxx User Population Estimates*, published by the Office of Public Health Support, Division of Program Statistics, Statistician and Director.

Dental Data Repository:

Additional dental encounter data is included in the User Population/Workload Data Mart, as needed for the Dental Data Repository and associated reporting.

Reporting:

- The TATONKA Database exists on the Smeagol server which is a separate server from the NDW Production Environment.
- Data in the TATONKA Database is refreshed for each User Population and Workload cycle.

- The Dental data owner requires NPIRS to produce reports on annual basis, which is usually during the March/April timeframe. Additional reporting is done on request from the Dental data owner. Current Dental data owner is Dr. George Chiarchiaro.

Data Parameters

The User Population/Workload (UP/WKLD) Data Mart contains all tables, procedures, and views needed for user population and workload reporting, including

- Registration data, as needed for user population reporting
- Encounter data, as needed for workload reporting
- Complete set of Reference tables for proper reporting

Reference tables are currently provided to the UP/WKLD data mart through DB2 Federation. The official single source control of the Reference tables resides in the TEMECULA database.

System Environment

For a detailed list of User Population/Workload related schemas and tables, see the following documents:

- *User Population Reporting Tables*
- *Workload Reporting Tables*
- *NDW Reference Tables*

User Population/Workload (UP/WKLD) Data Mart

Server:	SMEAGOL, 64bit
Database:	TATONKA, DB2
FTP Address	198.45.1.21
System Monitoring Tool(s):	IBM DB2 Query Patroller if necessary

NDW Production Data Mart

Server:	GIMLI, 64bit
Database:	WILDHRSE, DB2
FTP Address	198.45.1.9
System Monitoring Tool(s):	IBM DB2 Query Patroller if necessary

IHS Intranet – Web Reporting

Server:	ROHAN, 32bit
Database:	Microsoft SQL Server 2003
Windows Version:	Microsoft Windows Server 2003
Business Objects Version:	v11
Web Address	http://rohan.d1.na.ihs.gov
System Access:	Enterprise compliant to allow various environments to access the database and web, including ODBC, OLE and heterogeneous web connection tools.
System Monitoring Tool(s):	Microsoft auditing

Security

Access to the User Population/Workload (UP/WKLD) Data Mart will be restricted from outside customers. Access will be limited to NPIRS personnel for maintenance, data refreshing and report production.

Only authorized personnel will be allowed access to the database, and those authorized will have access to all data within the mart.

Only authorized users will be allowed access to the official IHS intranet reporting web site as directed by the Director, Division of Program Statistics, Office of Public Health Support (OPHS), Indian Health Service.

Security controls are commensurate with those in the production NDW database and adhere to IHS standards, as outlined in separate security documents.

Data Transfer Process

The data transfer process extracts selected encounter data from the NDW production tables in SANDIA, based on the specified criteria. A snapshot of the export information table (ADMIN.EXPORT_INFO) is taken at the start of each extract. This snapshot information is used by the extract process to control what exports are extracted. Only fully loaded exports are extracted, so that the Import Engine can continue to process while the extract process runs.

For logical diagrams of the User Population and Workload Data Transfer Processes, see Appendix A.

These are the types of extract processes used for the data transfer:

- 1) EL (Extract Load) extracts the tables inside Sandia into a flat file for subsequent load into another mart.
- 2) TEL (Transform Extract Load)
 - a) A variation of the EL process, TEL builds a flat file from a view or MQT (Materialized Query Table) that pre-exists by performing calculations or transformations with or without joins and/or OLAP (On Line Analytical Processing).
 - b) Builds a temporary table in SANDIA with calculations and transforms (with or without joins and/or OLAP), and then extracts these to another mart for load.
- 3) ETL (Extract Transform Load) is an extract process, then a transform at the receiving process, prior to or during the load. This is often done with third party software.
- 4) TTEL (Transform/Transform Extract Load) builds a temporary table in SANDIA with calculations or transforms (with or without joins and/or OLAP), updates the table, then extracts these to another mart for load; then potentially, there is further updating of the table in the mart.

User Population

Registration data is extracted to the NDW Production Data Mart's NDWMART schema and unduplicated across the entire population, using Integrity software. A separate utility then utilizes these results to determine the 'survivors' at both the area and national level. The results are then loaded to the User Population/Workload Data Mart (REG.USERPOP table) for reporting. Each user population extract process appends data to the REG_UNDUP table in the NDW Production Data Mart and the USERPOP table in the User Population/Workload Data Mart, using unique version numbers to identify the data set of each run. An archive process and/or compression techniques may be implemented at a later date, if required due to space limitations or reporting efficiency.

A log is maintained (REG_ETL_LOG table) to note the run date, visit date range, purpose, run type and version number of each unduplication run. Since multiple versions of userpop data may be stored in the mart, the version number will be used as part of the key into the User Population tables.

Tatonka contains additional legacy data for FY06 for those sites that were unable to transmit standard exports to the NDW. For FY07 forward it contains only NDW data.

Process Summary:

1. The userpop extract creates flat files:
 - The application is executed using certain criteria, such as:
 - Region - The normal option is all regions.
 - Date range - The date range applies only to the encounters that are utilized to determine whether the registration is active. Normal option = 3 years prior to specified end date. For example: if end date = 9/30/2005, the calculated start date would be 10/1/2002
 - Creates a log file documenting each run and version number.
 - Sets the active flag on each registration according to whether it had an appropriate workload-reportable visit within the time frame.
2. A separate Load process then loads files to the NDW Production Data Mart
3. The Integrity process is run.
4. Another process updates the NDW Production Data Mart with the Integrity id and current timestamp.
5. The Unduplication process is then run
6. Another extract process for the NDW Production Data Mart table creates another set of flat files
7. A Load Process updates/inserts the User Population/Workload Data Mart tables with the NDW Production Data Mart user population data.
8. An insert process updates the identical log file in the User Population/Workload Data Mart so reports can identify the appropriate version.
9. User Population data is available for reporting.

For detailed data transfer and unduplication specification, see the *Userpop Process FY06-Specification*, Version 2.0 (January 2007).

Workload

Not all visits are workload reportable. This is an important designation; therefore, the determination occurs at the warehouse level (set by the Import Engine at the time of the load to the NDW Production Data Warehouse). This eliminates the need for data marts to use their own (potentially distinct) logic, and the warehouse remains the single source of truth. For further information on what constitutes a workload-reportable encounter, see the *Workload Reportable Visits Technical Specification, Version 1.5* (February 2008).

Some data may be transformed or populated in order to make selection decisions or to standardize format during the extract, but production data is not modified.

- The NDW receives unfiltered data from the Fiscal Intermediary (FI), whereas the legacy NPIRS system received a filtered file. The workload database must be able to identify records that would have been received in the filtered file versus those that would not have been sent.
- The extract received into legacy NPIRS from the RPMS Patient Care Component (PCC) application had PCC edits applied that would restrict certain records from being sent. The NDW HL7 extract from the RPMS PCC system does not apply these edits, so the NDW receives records that previously would not have been sent.
- The determination has been made that only the following PCC edits will be applied to the NDW data after receipt.

PCC Edit#	Description	NDW Action if Edit Fails
H17	Discharge date cannot be prior to admission date.	Not extracted to User Population/Workload Data Mart for reporting.
H8	At least one diagnosis (purpose of visit) must be populated. Note: RPMS required this to be a primary diagnosis	Flagged as not workload reportable on load to NDW.
A6	At least one diagnosis (purpose of visit) must be populated. Note: RPMS allowed any diagnosis	Flagged as not workload reportable on load to NDW

Process Summary:

1. The workload extract creates flat files:
 - Attempts to calculate admission and discharge date if either null
 - Applies the H17 PCC edit. Encounters failing this edit are not extracted.
 - Selects via criteria such as current, non-deleted, workload-reportable encounters that are within the specified date range.
 - Default date range is all records. Run time specifications can be elected to narrow the date range.
 - Standard extract range is: from run date, determine prior October 1 date; start date is 2 years prior to that date. This range covers the current workload reporting period and 1 year prior for progress reports.
Example: run date is May 16, 2006. Start date is October 1, 2003. End date is run date.
 - For workload runs that coincide with User Population runs, the standard extract range is 3 years in order to extract all encounters that may have been utilized to activate registrations.
 - Set FI filter flag.
2. Flat files are then loaded into TATONKA.
3. Unduplication is performed in TATONKA via both update and MQT (Materialized Query Table) techniques.
4. Workload data is available for reporting.

For detailed data transfer specifications, see *UserPop/Workload Data Mart Technical Specification Workload ETL*, Version 1.7 (March 2007). For workload unduplication specifications, see *UP WL Data Mart (Tatonka) Unduplication - Encounter Specification*, Version 1.3 (September 2006).

Reporting

All standard reports are supported by views and stored procedures. Stored procedures support parameters used to adjust the reports.

Official reports reside on the NPIRS National Data Warehouse intranet web site:

<http://rohan.d1.na.ihs.gov>

The availability of published reports, once posted on the web site, will be independent of the User Population/Workload Data Mart refreshing or system maintenance.

- **User Population**

Reports are produced five times yearly (four drafts, one final), which also allows for quarterly GPRA reporting.

Official fiscal year end User Population reporting is aligned according to the approved *Schedule for Production of FY xxxx User Population Estimates*, published by the Office of Public Health Support, Division of Program Statistics, Statistician and Director.

- **Workload**

Reports are produced and posted to the NDW intranet web site on a monthly basis during the normal course of the fiscal year.

As the end of the fiscal year approaches, the frequency of posting workload reports can increase, according to the approved *Schedule for Production of FY xxxx User Population Estimates*, published by the Office of Public Health Support, Division of Program Statistics, Statistician and Director.

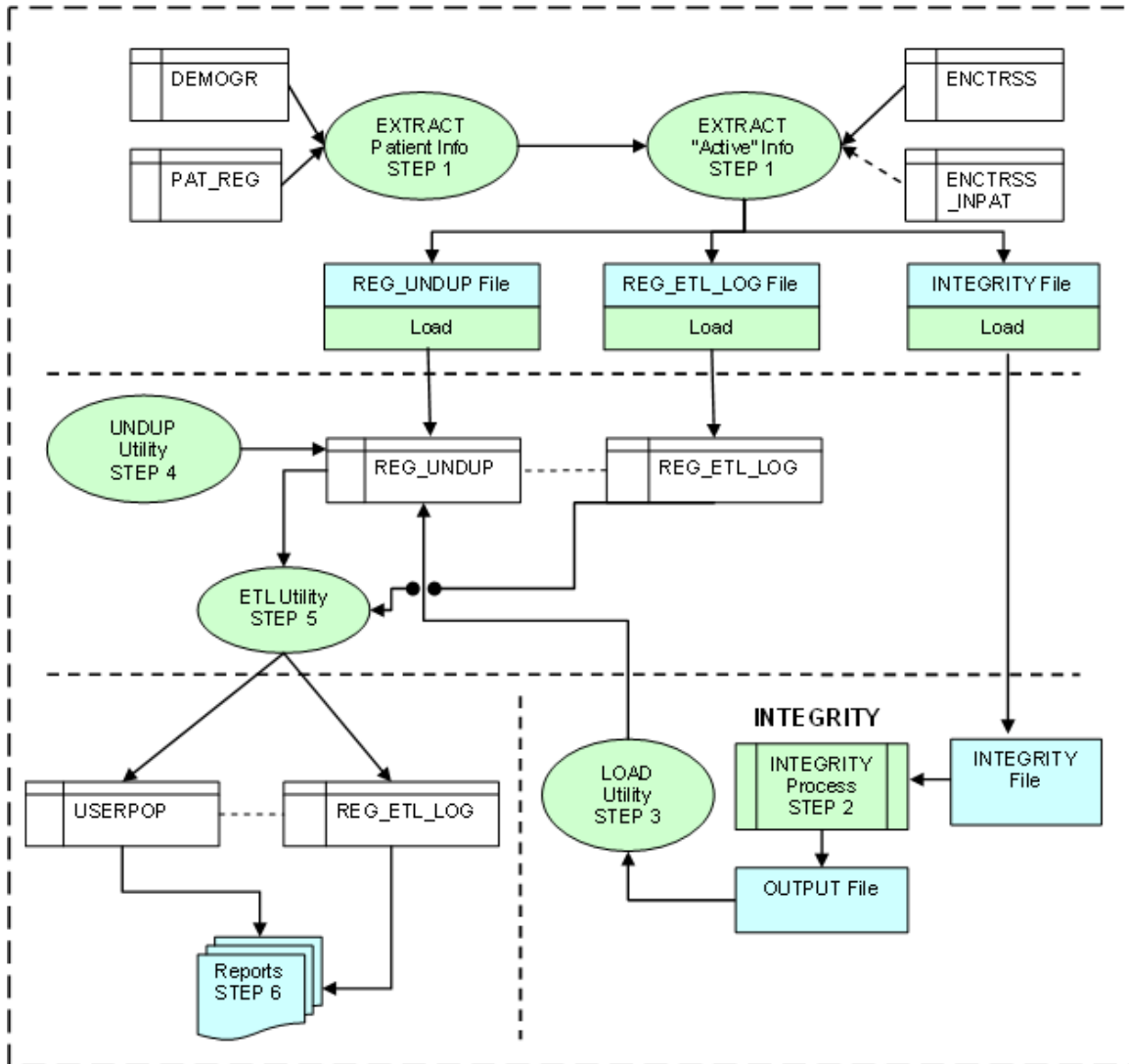
Archival

When reporting is complete for a fiscal year, all data pertinent to that year's user population and workload reports are archived. Details of the process may be found in the *UserPop/Workload Year-End Archive Process* (white paper) document.

Historical official reports remain available on the on the NPIRS National Data Warehouse intranet web site. The reports will be archived to a historical directory as required for space considerations.

Appendix A: Data Transfer Process Diagrams

User Population Process



Workload Process

