
National Data Warehouse

Export Tracking and Handling Process

Conceptual Design

Version 5.1

January 2005



Department of Health and
Human Services

Indian Health Service

Office of Information
Technology (OIT)

Contents

Overview	1
Background	1
Export Information.....	1
Fatal Errors.....	2
Serious and “Other” Errors.....	2
Assumptions	2
Export Tracking Process	3
Process Flow.....	3
Export Information.....	3
Fatal Errors.....	4
Serious Errors.....	4
“Other” Errors.....	5
Export Tracking Reporting	6
Serious Errors.....	6
“Other” Errors.....	6
Appendix A – Area Summary Report Sample	7
Appendix B – Acknowledgement Report Sample (sent via e-mail)	8
Appendix C – Post Load Report Sample (sent via e-mail)	9
Appendix D – Code Values	10
Appendix E – National Data Warehouse Tables	11



Version Control

Version	Date	Notes
1.0	October 2003	Creation date
2.0	June 2004	Publish date
2.0	September 2004	New doc template applied
3.0	October 2004	Separated the Export Tracker from the Data Quality Data Mart.
4.0	October 2004	Changed from requirements document to conceptual design.
5.0	November 2004	Final version submitted for customer approval.
5.1	January 2005	Includes customer revisions. Also added version number to footer.



Overview

The Indian Health Service is developing the National Data Warehouse (NDW) to store healthcare information at a central location in order to facilitate accessing information at the national level. One of the goals of this initiative is to improve the quality and integrity of the data that is collected and reported. For this purpose, the NDW has included a series of tables to track and facilitate reporting of the processing of the data exports.

Background

The NDW process for tracking the status of the data exports has three general categories:

- Export Information
- Fatal Errors
- Serious and “Other” Errors

Export Information

Export Information will be collected to track all data received by the National Data Warehouse (NDW). It is intended to store enough information to track the status of all exports. For example: the date the data was received at the warehouse, the date it was uploaded into the warehouse, the number of records received and processed, etc. This information is available for reporting to the customer.

Information specific to RPMS exports will also be stored and available for reporting to the customer, such as the number of records sent, Unique RPMS Database ID, PCC Export Log number, Total number of PCC Visits included in the export, etc.

Information such as number of records added, changed, and deleted will only be available after the data has been uploaded into the warehouse.

Area Summary reports will be made available to each area after the Initial Data Load is processed. These reports will include:

- the number of visits per Type and Category combination per Location of Encounter
- the number of visits per month and year based on visit date per Location of Encounter

(See Appendix A for sample Area Summary Report.)



Fatal Errors

Fatal Errors occur when the data that was transmitted to the National Data Warehouse (NDW) has somehow become corrupted, thereby making it impossible to load into the database. These types of errors will cause the entire transmission to be rejected, so they must be addressed immediately. The most likely resolution will be to contact the Integration Engine (IE) staff and/or the data originator to request the data be retransmitted.

Serious and “Other” Errors

Serious Errors are defined as records that do not contain enough information to determine its relationship to other data in the warehouse. For example, encounter records with no registration links, or chart information for a non-existent patient. These records will not be loaded into the NDW, but will be placed in an audit table with a description of why the record was rejected.

“Other” Errors occur at the field level but do not prevent the data from being loaded into the warehouse. There are currently two types of “Other” Errors: Transform Errors and Informational Errors. (Additional “Other” Errors may be added, but they will be processed and should be reported the same as Transform and Informational Errors.)

An example of a Transform Error is: if a date field contains only month and year the warehouse can not store that as a valid date so it must store it as a null (although the month and year is stored in a character field so the information is still available). An Informational Error would be, for example, if registration data is sent to the warehouse with no chart number. The data will still be stored but the customer will be informed that the NDW was not provided a chart number for that registration.

If Serious or “Other” Errors are found during the processing of the data export, an exception report will be e-mailed to the customer containing aggregate information about the errors, and the address for the proposed website where additional information may be obtained.

Assumptions

Fatal Errors will be handled as they occur on a case-by-case basis. The team responsible for loading the data into the warehouse (currently CNI/DataCom Sciences, Inc.) will also be responsible for reporting and resolving these errors.

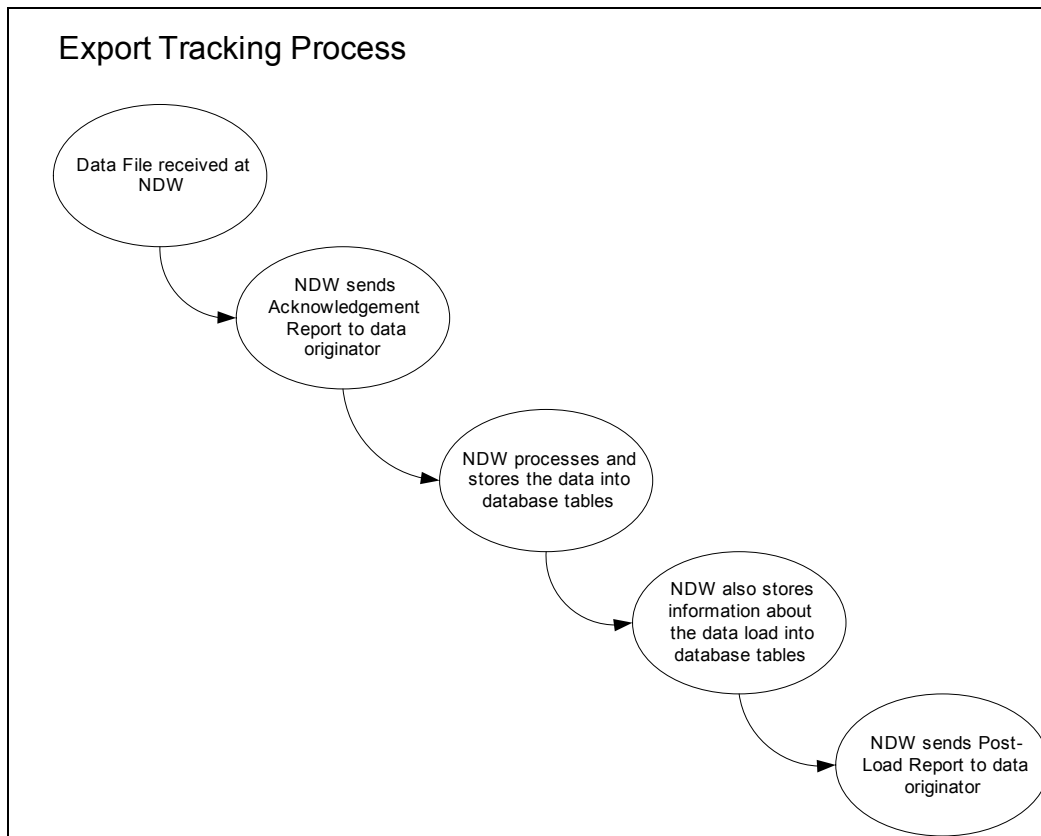
When a file is received at the warehouse and has passed the initial error check (i.e., no Fatal Errors were found) an e-mail will be sent to the data originator acknowledging receipt of the data. (*See Appendix B for sample Acknowledgement*). After the data has been processed and stored in the NDW database, an e-mail notification will be sent to the data originator indicating that the data has been loaded into the NDW tables. (*See Appendix C for sample Post Load Report*).



Export Tracking Process

Process Flow

Data is extracted from the local systems and sent to the NDW, which then processes and stores the data into tables within the database referred to as Target Tables. Similarly, the information about the data load is also stored in tables.



Export Information

Information about the data that is received by and subsequently loaded into the NDW will be stored in the warehouse table EXPORT_INFO. (See Appendix E for table layout.)

This table is updated during each step of the upload process and therefore all data may not be available at the time a report is generated.

Once the data transmission has been acknowledged, the following information will be available in the EXPORT_INFO table:

- Export ID
- IHS Area Code
- Name of File generated at the source (if available)

- Number of Records Transmitted from the source
- Date the exported data was generated at the source (presumably, the date the source transmitted the data to the Integration Engine.)
- Unique identifier to designate where the source system resides
- Name of Facility where the source system resides
- Begin and End Date the source used as selection criteria for the export
- Export Log Number assigned to the batch at the source (if available)
- Number of PCC Records (number of visits) transmitted
- Number of PCC Records not included in the export (skipped, errors, demos)
- Name of File generated by the Integration Engine
- Number of Records Transmitted from the Integration Engine
- Date the IE transmitted data to the NDW
- Error Category Code (Accepted or Rejected)
- Number of Records Received by the NDW
- Number of Header Records
- Number of Registration Records (number of patients)
- Number of Encounter Records (number of visits)
- Date file was received and acknowledged by the NDW
- Record Totals Match flag

The remaining information will not be inserted into the EXPORT_INFO table until after the data has been processed and loaded into the target tables:

- Number of New Registrations
- Number of Modified Registrations
- Number of Deleted Registrations
- Number of Rejected Registrations
- Number of New Encounters
- Number of Modified Encounters
- Number of Deleted Encounters
- Number of Rejected Encounters
- Date file was loaded into the NDW

Information about the rejected registration and encounter records will be stored in the LOAD_ERRORS table as describe in the Serious Errors section of this document.

Fatal Errors

Since Fatal Errors are handled during the data load into the National Data Warehouse (NDW), the processing will be described in the ETL documentation and will be handled by the NDW.

Serious Errors

As the data is loaded into the NDW, certain conditions may exist that will cause the warehouse to reject the data. For example, if we are unable to link a visit to existing patient registration information, that visit will not be added to the database.



When these errors occur, information will be saved in the warehouse table LOAD_ERRORS.
(See Appendix E for table layout.)

Each export will be assigned a unique identifier when it arrives at the NDW. This Export ID will be stored with the error information in the LOAD_ERRORS table so that errors can be reported by export.

The NDW contains a schema called ERROR. This is where the data that was “rejected” from the NDW is stored. This will allow us to join a record in the LOAD_ERRORS table to the registration or encounter that was rejected. It is anticipated that the data in the ERROR schema tables will remain until the issue is resolved and loaded into the target tables or the customer authorizes NDW to delete the rejected data.

The LOAD_ERRORS table will store different categories of errors as they are discovered throughout the load process. Rejected records will be assigned error category “R.” These are the only records that should be reported as Serious Errors. (See Appendix D for list of error category codes.)

“Other” Errors

Certain other conditions may occur where a particular field value is not able to be inserted into the warehouse (i.e. bad dates), or data appears to be “missing” (i.e. registration with no chart). Although the remainder of the data is stored in the warehouse, this may be information that the source of the data would like to know.

Information about these conditions is also saved in the warehouse table LOAD_ERRORS, (see Appendix E for table layout.) as describe for Serious Errors.

For additional information about these conditions, the NDW contains a schema called ARCHIVE. This is where the data is stored that is as close to “raw” as possible. The LOAD_ERRORS table can be joined to the registration or encounter data stored in the ARCHIVE schema. However, unlike the ERROR schema, the data in ARCHIVE will only be available for a limited time, the duration of which has not yet been determined.

Important note: the ARCHIVE schema is not used by the NDW as the permanent data archive. Extract files are stored on the disk as well as backed up to tape and stored off-site. In addition, the NDW database is backed up nightly and stored off-site. These off-site backups are currently marked for permanent retention. This means that ARCHIVE tables can easily be repopulated from more than one source at any time should the need arise.

The LOAD_ERRORS table will store different categories of errors as they are discovered throughout the load process. Transform Errors will be assigned error category “T” and Informational Errors will be “I.” Other error categories may be added to the cycle and should be included as an “Other” type of error.

Export Tracking Reporting

Serious Errors

Summary Information

Schema: ADMIN Table: LOAD_ERRORS

Identified as: Error_Category_CD equal to “R”

Reportable information: Export_ID, Uniq_Reg_Code, Uniq_Enctr_Code,
Error_Description

Detail Information

Schema: ERROR Table: All tables

All tables in the ERROR schema can be queried using the Export_ID and the Uniq_Reg_Code (when querying registration tables) or Uniq_Enctr_Code (when querying encounter tables).

“Other” Errors

Summary Information

Schema: ADMIN Table: LOAD_ERRORS

Identified as: Error_Category_CD not equal to “R”

Reportable information: Export_ID, Uniq_Reg_Code, Uniq_Enctr_Code,
Error_Description, Field_Name, Old_Value, New_Value

Detail Information

Schema: ARCHIVE Table: All tables

All tables in the ARCHIVE schema can be queried using the Export_ID and the Uniq_Reg_Code (when querying registration tables) or Uniq_Enctr_Code (when querying encounter tables).



Appendix A – Area Summary Report Sample

This is an example of what the report may look like.

Strawman sample of Encounter Summary Report				
Area: XXXXXXXXXXXXXXXX				
Service Type / Category				
Month/YR	Added	Changed	Deleted	Rejected
-----	-----	-----	-----	-----
X - XXXXXXXXXXXXXX / X - XXXXXXXXXXXXXXXX				
TOTAL	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
.
.
X - XXXXXXXXXXXXXX / X - XXXXXXXXXXXXXXXX				
TOTAL	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
.
.
X - XXXXXXXXXXXXXX / X - XXXXXXXXXXXXXXXX				
TOTAL	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
MMM YYYY	9999	9999	9999	9999
.
.



Appendix B – Acknowledgement Report Sample (sent via e-mail)

This is an example of what the report may look like.

Sample of NDW Data Transmission Log Report

* DATA WAREHOUSE *
NDW Data Transmission Log Report
Data Received Date: Aug 25, 2004

For more information please send an email to: ITSCDataWarehouse@ihs.gov

Export ID: 123
Export file name: 9999993040821090847.BDW
Static ASUFAC of exporting box: 999999
Beginning Date: Oct 01, 2000
Ending Date: Mar 31, 2001
Run Location: FACILITY NAME
Transmission Status: No errors detected
Total Number of Encounter Records Received: 0
Total Number of Registration Records Received: 99,999

Every file that arrives at the warehouse is assigned a unique Export ID number.



Appendix C – Post Load Report Sample (sent via e-mail)

This is an example of what the report may look like.

Sample of NDW Data Load Exception Report

* DATA WAREHOUSE *
 NDW Post Data Load Report
 Load Date: Aug 25, 2004

For more information please send an email to: ITSCDataWarehouse@ihs.gov

For error details, please visit: www.IndiansHealth.gov

A web site is being developed to provide additional information about the load exceptions.

Export ID: 123
 Export file name: 9999993640821090847.1
 Static ASUFAC of exporting box: 999999
 Beginning Date: Oct 01, 2000
 Ending Date: Mar 31, 2001
 Run Location: FACILITY NAME
 Load Status: Exceptions Detected

Some Export ID number as reported on the acknowledgement.

Total Number of Encounter Records Received: 0
 Total Number of Registration Records Received: 99,999

Encounters:	Add	Change	Delete	Rejected
	-----	-----	-----	-----
	99,999	99,999	99,999	99,999
Registrations:	Add	Change	Delete	Rejected
	-----	-----	-----	-----
	99,999	99,999	99,999	99,999

Error Description	Field Name	Count
*** TRANSFORMATIONS ***		
INVALID DATE	Date Moved To Community (character format)	4,567
INVALID DATE	Date of Death (character format)	53
INVALID DATE	Eligibility Start Date (character format)	1,234
STANDARDIZING GENDER	Gender	1



Appendix D – Code Values

Error Category Codes

Code	Value	Example
A	Accepted	Only used by EXPORT_INFO to indicate the export file was accepted and is being processed.
I	Informational	Only used by LOAD_ERRORS to indicate an Informational Error such as registration with no chart.
R	Rejected	Used by EXPORT_INFO to indicate the export file was rejected (Fatal Error). Used by LOAD_ERRORS to indicate a transaction was rejected (Serious Error).
T	Transformed	Only used by LOAD_ERRORS to indicate a Transform Error such as a bad date.

Appendix E – National Data Warehouse Tables

Insert table layouts from Data Dictionary for the following tables:

Schema	Table
ADMIN	EXPORT_INFO
ADMIN	LOAD_ERRORS
ARCHIVE	P000_ENCTRSS
ARCHIVE	P010_PROVIDER
ARCHIVE	P020_PROCEDURE
ARCHIVE	P025_HCPCS_PROC
ARCHIVE	P030_ADA_PROC
ARCHIVE	P03A_TEETH
ARCHIVE	P040_DX
ARCHIVE	P050_EXAM
ARCHIVE	P060_PAT_EDUCATION
ARCHIVE	P070_IMMUN
ARCHIVE	P080_LAB_TEST
ARCHIVE	P090_CLIN_MEAS
ARCHIVE	P100_PAT_SKIN_TEST
ARCHIVE	P110_MEDICATION
ARCHIVE	P120_HEALTH_FACTOR
ARCHIVE	PHDR
ARCHIVE	R000_PAT_REG
ARCHIVE	R010_DEMOGR
ARCHIVE	R020_ALIAS
ARCHIVE	R030_CHART
ARCHIVE	R040_INSUR_ELIG

