

# Status of the MHDO Health Data Warehouse Master Indexes

---

## Overview

MHDO's vision of the Health Data Warehouse project originally called for the creation of three master indexes: payer, patient, and provider. In its call for proposals, they stated that these indexes would "create a shared utility that will provide value for multiple entities through the state." These indexes would be used across claim, hospital, and other data streams to provide "consistent, accurate and historical demographic data" on patients, providers, and payers.

During the past two years, MHDO and its contractor HSRI have made major strides towards meeting these requirements. This document outlines what has been done in relation to each of the three indexes and what remains to be done to fully achieve the goals articulated above.

## Master Payer Index

The Master Payer Index is intended to provide "consistent, accurate, historical, and current demographic data on the payers...reported across the claims, inpatient/outpatient, and other data streams. Each payer was to appear only once across all streams in this index."

Currently, every payer that submits claims data is assigned an MHDO Assigned Code. This unique identifier and the payer name are stored on the Payer table, along with the date of addition. The MHDO Assigned Code and Payer name are distributed to data users as a part of the Data Release process. In addition, MHDO and HSRI have created a report that documents the activation and deactivation dates of any new payers or payers that no longer meet the submission threshold.

## Steps to Achieve Original Goals

While much of the information articulated in the original call for proposals is being tracked, this is not currently aggregated in a single data structure. In order to achieve a true Payer Index, the current Payer table should be enhanced to include data elements such as activation/deactivation dates, and data start/end dates. Tracking of historical name changes (if any) could also be considered.

Claim data is currently the only data stream for which we receive payer information directly.

## Possible Enhancements outside the Scope of Original Proposal

It is possible that data users could benefit from having a Payer Directory. This directory would go beyond the "demographic" information offered by the index and provide value-added information about each entity. This information could include details about the types of coverage offered, information about entity relationships, and other similar types of data.

It is also possible that other data streams, such as hospital data, could be enhanced through the addition of payer information. Doing this would be contingent on being able to reliably associate inpatient/outpatient data with claims data. This might enable analyses of treatment patterns by payer, etc.

## **Master Patient Index**

The Master Patient Index is intended to provide “consistent, historical, and current demographic data on the patients reported across the claims, inpatient/outpatient, and other data streams.” Each patient in the index would receive a single unique identifier across all the streams.”

Currently, the MHDO calculates a unique member ID based upon the member SSN, the subscriber SSN, and/or the contract number on the claim. A project was recently completed that performed partial de-duplication of historical member IDs. However, it is known that some duplication of IDs (that is, situation where one individual has more than one member ID) still exist due to ambiguities in the data.

## **Steps to Achieve Original Goals**

While the MHDO currently provides a unique identifier for each member, to the extent it is able, it does not currently maintain a table that provides the “current and historical” demographic data in one data structure. This information is available on the claims tables, but has not been specifically aggregated. The MHDO also does not currently create member IDs for the hospital data. This makes it difficult to begin the process of linking these data to claims. In order to achieve a true patient index, a single data structure should be created to store member information from both the claims and hospital data.

In order to improve the utility of the member ID, a method of “disambiguation” should also be created that would make use of name information to attempt to resolve situations where other fields provide conflicting information. This information would be used to generate a list of “candidate matches” that could be used to manually indicate two entities are one individual (join) or one entity is actually two individuals (split). A record of these split and join decisions should be captured in the form of metadata associated with the records. Currently, while manual splits and joins are documented, this documentation is not made in such a way that it is easily associated with the underlying data rows.

## **Possible Enhancements outside the Scope of Original Proposal**

The detection of potential joins and splits could be improved through the addition of probabilistic matching methods using a tool such as Insight or Mirth Match. Also, the data submission rule could be enhanced to require that payers submit enhanced demographic information about patients in a separate “demographics file” or in the eligibility records. This would enhance the ability to accurately identify individuals within the claims data and also improve the potential for making linkages to other data sources, such as hospital and clinical data.

## Master Provider Index

The Master Provider Index is intended to provide “consistent, accurate, historical, and current demographic data on the medical providers reported across the claims, inpatient/outpatient, and other streams.” This index was required to include the National Provider Identifier (NPI) of the provider.”

Currently the MHDO APCD maintains a Provider Master File that contains a unique identifier (DPCID), provider name and a small amount of additional information. This is linked to the Provider Detail File that provides a unique identifier (PRVIDN) and relevant provider information that appears on the claim. Both the detail table and the master file contain a field for NPI, however, this field has only started to be widely populated in the last few years. Work is underway confirming the NPIs that exist in the master file and adding as many missing ones as is possible. Extensive work has also been done to manually split or join historical detail records to the appropriate master file row.

The hospital data currently identifies providers through the use of the NPI. Thus, a linkage can be made between hospital data and claims data using the Provider Master File, as long as the physician indicated in the hospital data also appears on the claim.

The combination of the master file, the freely available National Provider registry (NPPES registry), and the provider detail table provides a large amount of information on providers. The table below indicates the fields that are currently available (all the rows that have a checkmark in the N or C columns).

## Steps to Achieve Original Goals

Through the use of the NPI as the primary provider identifier, the MHDO data warehouse allows data users to access current demographic data on medical providers. This information is filled in from what was provided on the claim in situations (primarily historical in nature) where no NPI is available. While the NPPES registry provides fields to record other names associated with an individual or organization, there is no way to see the history of such name changes. HSRI has already created data structures to allow information such as name and license history from the NPPES to be stored. The Data Warehouse provider information should be transitioned from its current data structure to this new one and periodically updated from the NPPES to capture periodic changes. Also, since it appears that older NPIs can “drop off” the NPPES registry, the current state of certain fields should be stored in our data structure to ensure that future users of the data have access to this information.

While the hospital data currently includes provider NPIs, this process should be enhanced to also associate these data with provider master file/provider index entries. This will make it easier to link these data with claims data.

The release format of the provider master and detail files that accompany the claims data should also be evaluated to determine what, if any, new provider fields should be included. Similarly, an evaluation should be made of whether to begin including provider information files to accompany the hospital data.

**Table 1: Potential Provider Data Elements by Level and Possible Data Sources**

**Primary Data Sources:**

- B** = Data is available from boards of licensure
- N** = Data is available from NPI Registry
- C** = Data available from claims submitted to payers
- R** = Payers' provider rosters

**Key to Symbols:**

- ✓ = Generally available from this source
- P** = Available for some providers, practices or sites, but not all
- ✗ = Data not available from any automated source

B	N	C	R	Individual Providers
<b>Provider ID (computer generated)</b>				
✓	✓		✓	Provider's Name
✓	✓		✓	Credentials (e.g., M.D., D.O., etc.)
✓	✓		✓	Gender
✓	✓			Maine Medical License #
✓				Maine License Status
✓				Maine License Date Issued
✓				Maine License Exp. Date
✓				Active or Inactive License
		✓		MaineCare ID Number
		✓		Medicare Provider Number
✓	✓		✓	NPI Number
✓	✓		✓	Provider specialties
			✗	Email Address
P	✓		✓	Primary Practice Site

B	N	C	R	Practice Sites
<b>Site ID (computer generated)</b>				
	P	P	✓	Practice Site Legal Name
	P		✓	Practice Site Other Name
			✓	Practice Site Type*
	P		✓	Practice Site Mailing Address
	P	P	✓	Physical Address
	P		✓	Site Phone Number
	P			Site Fax Number
			✗	Website URL
	P		✓	National Provider Identifier (NPI)
		P		Federal Tax Identification No. (TIN)
		✓		Site MaineCare ID
		✓		Site Medicare ID
	P	✓	✓	Practice specialties
	P	✓	✓	Practice Type (e.g. "Primary care")
	P			✗ Contact Person or Liaison:
	P			✗ Name
	P			✗ Title
				✗ Email
	P			✗ Phone

B	N	C	R	Practice Organizations
<b>Practice Organization ID</b>				
		✓	✓	Practice Org. Name
		✓	✓	Practice Org. Mailing Address
			✗	Practice Org. Email
			✗	Practice Org. Phone Number
			✗	Practice Org. Website URL
			✗	Administrative Leader (e.g. CEO)
			✗	Name
			✗	Title
			✗	Email
			✗	Organization Main Contact
			✗	Name
			✗	Title
			✗	Email
			✗	Phone

B	N	C	R	Broader Entities
<b>Broader Entity ID</b>				
				✗ Broader Entity Name
				✗ Mailing Address
				✗ Email
				✗ Phone Number
				✗ Website URL
				✗ Broader Entity Type†
				✗ Administrative leader (e.g. CEO)
				✗ Name
				✗ Title
				✗ Email
				✗ Administrative Office Main Contact
				✗ Name
				✗ Title
				✗ Email
				✗ Phone

\* e.g., solo practice, group practice, FQHC, RHC, hospital, VA, etc.

† e.g., integrated healthcare network, accountable care organization, provider network, etc.

## **Possible Enhancements outside the Scope of Original Proposal**

The provider index information discussed above allows data users to identify providers in claims and hospital data. However, it is not a true provider directory. The table above outlines data fields that have been identified as being potentially useful to data users. In addition to the basic demographic information already maintained, a directory could document the relationship between individual providers, practice organizations, practice sites, and broader entities. As shown above, much of this information is not available directly from the claims or hospital data. It would need to be obtained from other sources.

Identifying reliable data sources for the above information would allow the MHDO to develop a true provider directory. It would also, potentially, provide the opportunity for detecting changes in provider demographic information that had not yet appeared in the NPPES registry. Thus, this directory could be useful for a wide variety of entities.

It also might be useful to data users to enhance the provider index/provider directory with information such as quality metrics, education level, etc. This information may be obtainable from third party sources.