Proposal for: New Hospital Data Versioning System, Structure for Release Notes and Control Totals

September 26, 2017

# Overview

MHDO is proposing a versioning system that would allow changes to the inpatient and outpatient hospital data to be more effectively tracked and communicated to the data users. This versioning system would allow sets of records associated with previous data releases to be easily identified and retrieved even if subsequent processing has replaced or otherwise modified some of these records.

Each release or modification of data on the inpatient or outpatient data will be associated with a distinct version number (the inpatient and outpatient version numbers will be independent of each other). So, for instance, all of the hospital data that has currently been released in the new master table formatted inpatient file would be version 1. If we then had a facility resubmit data from 2013-2016, that new data would be version 2 and so forth. Then, if we removed some 2015 records for another facility, that change would be version 3.

Using the information recorded on the master tables, we would be able to pull the set of 2015 records that were distributed for version 1. These data wouldn't include any of the new data added as a part of version 2 and it would include the data that was removed as a part of version 3.

# Tracking of Version Control Numbers

Both the inpatient and outpatient master tables would have an associated version tracking table. This table would provide a brief user-facing summary of the changes associated with the version number: e.g., version 3 – removed 2015 records from facility X; version 4 – Initial release of 2017 Q1 data. This version tracking table will be made available on the MHDO website as a part of the data release documentation.

Release notes will continue to be associated with a year of data (e.g., Inpatient 2016). However, each iteration will also be associated with the latest version number released. So, for instance, the current set of Inpatient 2016 release notes would be tagged with version 1. If, in the future, new 2016 Inpatient data were added, those changes would be associated with a higher version number (let's say 5). If the changes were material enough to warrant it, a new set of release notes might be issued, in which case the new version of the notes would be tagged with a version 5.

The highest version number of included data would also be included in the file names that are sent to data users when servicing data requests. So, for instance, a new data user might get Inpatient 2016 data tagged with version 1. At a later date, they would be able to see on the MHDO website that version 5 included changes to 2016 Inpatient data. If version 5 included new data, they could request that they be sent just the new records; if the version removed data, they could request to be provided the IDNs of any dropped rows.

**SDUG Feedback:**

# New Structure for Release Notes

The goal is to improve the end users access to the MHDO’s data release notes. Based on our discussions with the data users the proposal is that the current Word-based release notes would be replaced by a new set of table-driven release notes. These release notes would be included as a table that would be packaged with each data release. The table would include all release items from the current or previous release versions with a data year value that was included in the release (release note items tend to apply to a single year of data).

**Figure 1: Outpatient Release Note Table Sample**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Note ID** | **Release Version** | **Note Text** | **Start Data Year** | **End Data Year** |
| 1 | 1 | This is the first release where all data were submitted through the new MHDO Hospital Portal. One of the primary goals of the new Portal is improving data quality by monitoring and validating data at the point of submission. The new system utilizes approximately 300 rules and thresholds to validate the content and completeness of the outpatient data. | 2016 |  |
| 2 | 1 | This release includes encounter data for: 39 Hospitals, 850 Physician Practices (primary care, specialty care and clinics), 105 Other, which includes labs, imaging centers, therapeutic centers | 2016 |  |
| 3 | 1 | Effective with the implementation of MHDO’s Rule Chapter 120 –Release of Data to the Public, by default, data releases will include patient county and age. Patient city, Zip code and date of birth will require special justification for inclusion. | 2016 |  |
| 4 | 2 | APR and MS DRG values have been recalculated using version 2017.1.1 of the 3M Core Grouping software, which includes v34 of both the APR-DRG and the MS-DRG. Record with discharge dates from 1/1/2016 - 9/30/2016 had values calculated using v33 of these groupers and records with discharge dates on or after 10/1/2016 had values calculated using v34. In the previous release of these data, all DRG values were calculated using v33 due to software availability. | 2016 |  |
| 5 | 3 | Populated field OP7104\_PrincipalDiagnosis from field OP7404\_OtherDiagnosisCode1 when OP7104 is NULL and OP7404 is not NULL. | 2015 | 2016 |

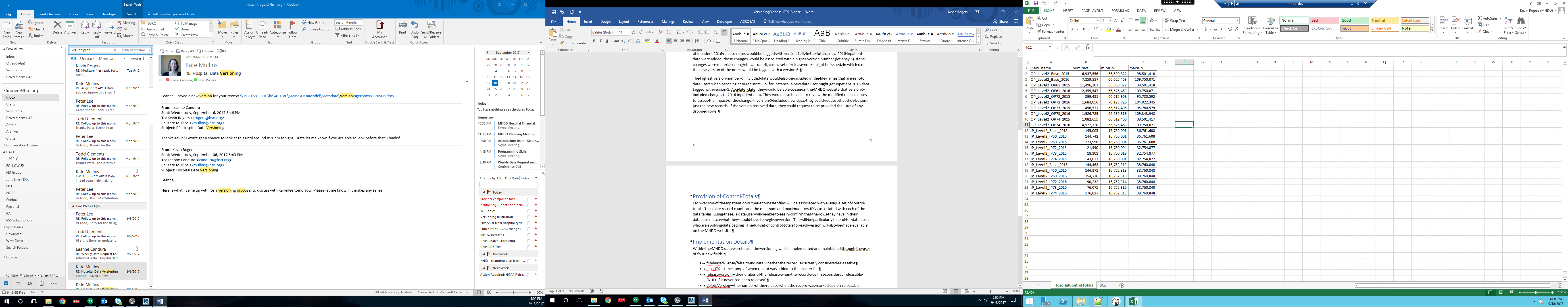
The above table shows some sample release note items from the Outpatient Release Note Table. Note that item #4 only applies to release versions 2 and would only be included in data releases that got 2016 data. Item #5 only applies to release version 3 and later; it would be included in data releases that got either 2015 or 2016 data.

**SDUG Feedback:**

# Provision of Control Totals

Each version of the inpatient or outpatient master files will be associated with a unique set of control totals. These are record counts and the minimum and maximum row IDNs associated with each of the data tables. Using these, a data user will be able to easily confirm that the rows they have in their database match what they should have for a given version. This will be particularly helpful for data users who are applying data patches. The full set of control totals for each version will also be made available on the MHDO website.

**Figure 2: Inpatient and Outpatient Control Totals – 2015, 2016**



The above figure shows the current release version's control totals for the Inpatient and Outpatient data for 2015 and 2016.

**SDUG Feedback:**

# Implementation Details

Within the MHDO data warehouse, the versioning will be implemented and maintained through the use of four new fields:

* flReleased – true/false to indicate whether the record is currently considered releasable
* insertTS – timestamp of when record was added to the master file
* releaseVersion – the number of the release when the record was first considered releasable (NULL if it never has been released)
* deleteVersion – the number of the release when the record was marked as non-releasable (NULL if it has never been removed)

In order to get the current set of releasable records, queries will be set up to pull all rows where flReleased = true. In order to see all of the data associated with a previous version, queries will be set up to pull all rows where the releaseVersion is less than or equal to the version of interest and the deleteVersion if either NULL or greater than the release of interest. In order to see all unreleased data, queries will be set up to pull all rows with an insertTS greater than the highest released insertTS.

The release number associated with a given data file would be the maximum releaseVersion number of any record contained within it or the maximum deleteVersion of any record that would have been included if it hadn't been included, whichever is higher. So, for instance, if a user was getting 2013-2015 outpatient data and all of the 2013 records were version 1, that is the version number that would be associated with the 2013 files. If the 2014 data had rows from versions 1 and 3, all of the 2014 data files would be associated with version 3. If the 2015 records had versions 1 and 3 and version 4 included 2015 records that were deleted, then the 2015 data files would be associated with version 5.

The goal is that the version number associated with a data file indicates the set of control totals that should be consulted to validate the record counts of the various tables. The version number also indicates which set of release notes are applicable to a set of released data—that is, release notes for the given year's data with a release version equal to or less than what they received.

MHDO plans to implement this proposal with the release of the Q1 2017 inpatient data.

**SDUG Feedback:**