MHDO Data Dictionary Suggestions and Updates

October 18, 2019

This document contains a list of suggestions from the SDUG for the MHDO Data Dictionary as well as proposed implementation timelines.

| **#** | **Suggestion** | **Resolution Plan** | **Ease of Implementation** | **Implementation Timeline** |
| --- | --- | --- | --- | --- |
| 1 | Add Fixed Storage Width and Data Type(s) to Configure Table Columns so they can be included in download. | Implement as suggested | Easy | MTNA to confirm this information is currently available in the data structure propose timeline interface updates |
| 2 | Store a recommended storage format with each item (e.g, date, varchar, currency) and display this attribute on the main screen columns so people can download or print the information. | Implement as suggested | Easy | MTNA to confirm this information is currently available in the data structure propose timeline interface updates |
| 3 | Add an indicator as to whether the field is calculated or a direct input. Refer to business rules documentation for details of calculations. | Implement as suggested | Medium | HSRI will provide information to MTNA by 12/2  MTNA to propose timeline for adding this to the data structure and interface |
| 4 | Improve search functionality. (Ex: Illustrating the importance of how important it is to get search functionality right try going to the data dictionary and searching for “SSN”. You will see that no results are returned. Now search for “MEMSSN”. You will see all of the results relating to members social security number. This little example illustrates a defect in the search functionality of the data dictionary since I think it is reasonable that the search term “ssn” should match any variable containing the substring “SSN” such as “MC010\_MEMSSN”.) | Discuss | Potentially easy | Discuss with MTNA |
| **5** | **Code Sets**  Adding code sets for example is something that should be fairly easy aside from the licensing issue since there already are several code sets. One place to quickly get information about almost any code set is: <http://athena.ohdsi.org/search-terms/terms>  One frustration people have with the data dictionary is that, rather than showing values and their associated meanings, the data dictionary often just has a "view classification -- external resources" link to <http://www.nubc.org/subscriber/index.dhtml>,  which as the name implies, you need a paid subscription to in order to view the codes.  If you just google the field, you can often find the code list on a non-paid site (often an insurance company, but sometimes in CMS memos).  I have an old hard-copy set of some code lists I got in 2004 or so. | We are unable to publish code sets which are obtained under a paid license. We will do a comprehensive review of the information currently in the dictionary and make updates to ensure users can easily be routed to the source. | Medium | HSRI will review all classification information and provide updated information to MNTA by 12/2  MTNA to propose timeline for interface updates |
| **6** | More metadata about each variable.   * How often is this variable missing? * What can this variable be used for? How has it been used? * Are there certain values that users should be aware of such as an empty string ‘’ or blank space ‘ ‘? * What is the format of the variable? For example, revenue codes are stored in the database as a fixed storage width of 10. However, these codes are actually represented in the data as 2 digit (if leading zeros are excluded) or 4 digit codes (if leading zeros included). It would also be helpful for someone looking at the page to know that sometimes leading zeros are included and sometimes they are not. (It would be even better to standardize the code representation to either 2 or 4 digits.) * What is the range of possible values for this data element? Are there dates from 1900 for example and how should the user interpret a date from 1900? | Discuss, some may be more difficult than others. |  |  |
| **7** | Variable descriptions should be improved and written by someone who is familiar with how the data element is used in practice. For example, look at the description for MC054\_REV (revenue code). It does not mean much to a data user. Compare the description of revenue code on the MHDO data dictionary to the description here: <http://www.medicalbillinganswers.com/revenuecodes.html>.  This latter description is much more helpful.  Sometimes a description of a field is sometimes just a repetition of the field name.  For example, the description for the IP2007\_PointOrigin field is "Point of origin for admission or visit." | HSRI will review descriptions | Medium | HSRI will provide information to MTNA by 12/2 |
| 8 | Also, the online data dictionary is really for the current data structure, which only began being used a few years ago. | None. The data dictionary is meant to reflect the currently available data structures. | Discuss | None |
| 9 | You can see the type of table I was thinking would be helpful for organizing documentation links if you go to <http://www.nber.org/data/vital-statistics-natality-data.html> and scroll down to the United States -- Data & Documentation table.  The columns would be different for MHDO ... maybe Year, Layout, Data Dictionary, Release Notes, and Supplemental Materials ... with one table for Inpatient, one for Outpatient, and one for Claims. | Review data user documentation currently made available during the October data user group meeting. | Discuss | October 30, 2019 |