The MNU Well Identifier Model

MNU stands for Minnesota Unique. The MNU well identifier model describes the uniqueness characteristics of well identifiers recorded in table o1id of the OWI database. That table corresponds to table c5id in the CWI database. The o1id table has the same fields as the c5id table, plus two additional fields for classifying the identifiers: integer field MNU, and integer or boolean field sMNU. The MNU number is used to identify which identifiers are considered to be MNU identifiers, and to implement uniqueness constraints on them. The sMNU number selects a single MNU number for each well as the default value to use when a single value must be selected without other guidance. The MNU number is also used in building identifier relationships needed for sorting out complicated real-world use cases.

MNU identifiers, Unique Well Numbers, and Well Record Numbers

Identifiers fall into two broad classes: those that have a uniqueness relationship with wells defined for the entire state of Minnesota, and those that do not. The first class are given non-zero MNU numbers: MNU $\in [1, 2, 3, \ldots]$. Examples of identifiers in this class include Minnesota Unique Well Numbers, Well and Boring Construction Record Numbers, and Well and Boring Sealing Record Numbers.

Identifiers that do not have have a uniqueness relationship with wells defined for the entire state of Minnesota are assigned MNU = 0. Examples of identifiers in this class that are found in table c5id include local well identifiers, county assigned permit numbers, and parcel identification numbers. These kinds of identifiers are not part of the MNU model, and are not discussed further.

MNU Identifiers have a close affinity to Minnesota Unique Well Numbers, but are defined differently. Minnesota Unique Well Numbers are defined as the value held in the wellid field of table c5ix. They are the numerical equivalents of the text values in the RELATEID field and the UNIQUE_NO field of the same table. Because the wellid field is a relational identifier field for the CWI database, it is required to have a 1:1-uniqueness relationship with wellids. The primary limitation of this is that redundant records cannot be de-duplicated except by deleting one of the redundant records, and also its wellid.

The 1:1 uniqueness of the wellid results in very simple identifier model. It is easy to work with, and it accords with the original concept for assigning Minnesota Unique Well Numbers. Over the years it has been proven that identifiers *can* be assigned uniquely and permanently. On the other hand, it has been proven that one *cannot* prevent multiple Minnesota Unique Well Numbers from being assigned to the same well. As a consequence, many Minnesota Unique Well Numbers have had to be deleted. An identifier whose definition can be deleted is impermanent, and is not suitable for sharing in the wide world. *Impermanence* is an essential property of Minnesota Unique Well Numbers that makes them unreliable for sharing outside of the CWI database.

In addition to deleting Minnesota Unique Well Numbers, altering the definitions of Minnesota Unique Well Numbers has also been allowed. This happens occasionally when well construction or sealing Record numbers are used as Unique Well Numbers, and single records are allowed to describe to multiple wells. It is not really necessary to alter the definition of the original identifier in these situations, and the practice should not be allowed. *Impermanence* of Minnesota Unique Well Numbers is a necessary outcome if altering their definitions is allowed.

Using exclusively the wellid to identify wells is ideal because they are integers and because the CWI database guarantees its validity and its uniqueness. The users who most realize this benefit are the owners and maintainers of the CWI database. For other users, including well owners, well drillers, local governments, consultants, and others, the impermanence of the Minnesota Unique Well Numbers that they have recorded in their own records can be a significant burden.

Permanence of MNU identifiers

The reason for introducing the concept of MNU identifiers, is to have a system of well identifiers that is reliable for sharing anywhere and by anyone in the state of Minnesota. The MNU identifier model builds upon the Minnesota Unique Well Number model by adding the critical property of *Permanence*. Permanently reliable well identifiers are suitable for use both inside and outside of the CWI database, and are suitable for use in archival quality records, whether paper or electronic.

The identifiers included into the MNU model are:

- Minnesota Unique Well Numbers, both current and historic
- Minnesota Well and Boring Construction Record Numbers
- Minnesota Well and Boring Sealing Record Numbers

This list is limited for practical reasons, but in theory can be expanded.

Because it includes all historically assigned Minnesota Unique Well Numbers, the MNU model must accept identifiers from the W-number series, and the identifiers must be defined as text values rather than an integers. This means that it can naturally accept H-number identifiers from Sealing Records. The H-number identifiers are included because they are extremely important to well owners and well drillers.

Because the MNU model includes all historically assigned Minnesota Unique Well Numbers, it must handle the cases where a single well has been assigned multiple Unique Well Numbers over time. This means that it will naturally accept both Construction and Sealing Record Numbers from the same well.

Because the MNU model includes all Construction and Sealing Record Numbers, it should attempt to handle the added complexity of those numbering systems. In particular, individual Construction Record Numbers and Sealing Record Numbers are allowed to refer to sets of wells (one record describing multiple wells). Adding this complexity to the MNU system represents trading of simplicity in definition for simplicity in application. The Minnesota Unique Well Number model is very simple in permitting only a single integer valued identifier per well, but handling W-numbers, H-numbers, and multiple-well Record numbers is difficult. The forced simplicity of the system provides little guidance to database programmers, to data maintainers, or to data users when more complicated identifier relations are encountered. The MNU identifier model attempts to internalize certain real world complexities, to make them normal cases rather than exceptional cases, to give clear guidance to programmers and maintainers, and provide an experience for data users that is as simple and easy to understand as possible. The MNU identifier model enables the CWI database to enforce uniqueness constraints on all MNU identifiers, resulting in higher quality data. The MNU identifier model gives a very brief set of rules for describing complex identifier relationships that should handle nearly every case known.

MNU relationship types and type numbering

The MNU identifier model utilizes the o1id table for storing all MNU identifiers. The o1id table is like the c5id table, but adds two integer fields: MNU and sMNU. The MNU field is set to a positive integer (the MNU relationship type number) for identifiers that are considered MNU identifiers, and to 0 for other identifiers in the o1id table. The sMNU identifier is set to 1 for exactly one identifier per wellid, and to 0 for all others. The choice of which identifiers are set to sMNU=1 is fundamentally arbitrary and is subject to change.

The MNU model assigns a relationship type number of MNU=1 to identifiers that are assigned uniquely and permanently to exactly one well. These includes the vast majority of existing Minnesota Unique Well Numbers, Minnesota Well and Boring Construction Record Numbers, and Minnesota Well and Boring Sealing Record Numbers.

The MNU model assigns a relationship type number of MNU=2 or MNU=3 to identifiers that are assigned uniquely and permanently to sets of wells. This includes Minnesota Well and Boring Construction Record Numbers and Minnesota Well and Boring Sealing Record Numbers that describe multiple borings.

The individual wells described on multiple boring records are sometimes also assigned individual Minnesota Unique Well Numbers. The identifiers for the individual wells are assigned MNU=1. When the individual wells are *not* assigned their own wellids, then the identifier describing the set is assigned MNU=2. When the individual wells *are* assigned their own wellids, then the identifier describing the set is assigned MNU=3. That distinction helps to avoid counting wells twice in certain analyses.

Programmers and data maintainers need to have a good understanding of how to assign and use the different MNU numbers. Data consumers generally need to distinguish only between MNU and non-MNU identifiers (MNU=0 or MNU$\in (1, 2)$). Illustrations of these relationships and references to real-world examples are provided in Appendix A.

Cross referencing MNU identifiers in sets of wells

When a Well Record identifies a set of wells with a single MNU identifier (the record number), and the individual wells in the set are also identified with their own individual MNU identifiers, then there should be means for

documenting and navigating those relationships. It would be possible to construct a relationships table for that purpose, and to build appropriate relationships to the o1id table and appropriate data integrity constraints - that approach is not taken here, but should be considered.

It is also possible to represent the relationships inside of the o1id table as follows: The individual well identifiers are entered linked to their own individual wellids, and assigned MNU=1 The identifier for the set of wells is entered once linked to the wellid describing the set, and assigned MNU=3. The identifier for the set is then entered again for each individual wellid in the set, and assigned MNU=4; these entries are the cross references.

Uniqueness constraints for the uniqueness of MNU identifiers are applied only to records with MNU $\in (1, 2, 3)$. Uniqueness constraints for cross reference entries are applied only to records with MNU=4.

Views of the o1id table that simplify data interpretations are constructed by including or excluding certain MNU types.

MNU default identifiers

The concept of a default identifier is simply to flag a single MNU identifier for each wellid that can be used in data reporting where the report requires selecting a single identifier for each well. The MNU identifier rules try to treat all MNU identifiers as co-equal so that they will all be equally functional for every use; therefore the choice of a default identifier should be fundamentally arbitrary. But it is advantageous if the choice is both automatic and easily predictable, so the following arbitrary rule is introduced: *The default identifier is the lowest valued alpha-numeric MNU identifier, with numbers sorting before letters.* With this rule, the default identifier will normally be the same as the identifier currently selected as the Minnesota Unique Well Number.

It must be emphasized that *the choice of default identifiers cannot be guaranteed permanent*, for the same reason that Minnesota Unique Well Numbers cannot.

A general scheme for querying CWI for data using a list of MNU identifiers would be to report back both the MNU identifiers used in the query and the identifiers currently flagged as default.

Integer valued identifiers

The MNU identifiers are defined as text, and allow both alpha and numeric characters. But there are many data manipulation processes that are more difficult when alpha valued identifiers are mixed in with numeric valued identifiers. The three most important are 1) use as a relational identifier within a database, where integers are preferred, 2) importing or exporting well data in spreadsheet programs, where default formatting of numbers as text can cause issues, and 3) variations in the use of leading zeros when formatting the number part of an identifier as text. CWI defines the Unique Well Number to be equivalent to the wellid, which is the relational identifier, and it declares the wellid as an integer.

MNU identifiers do not have to be integers because they are not required to be equal to the relational identifiers; a well could be added to the database having *only* an H-number or *only* a W-number as its MNU identifier. But the benefit to data manipulation outside of the database may still justify requiring that an integer valued MNU identifier be created for each well. In that case, it would be logical to create them equal to the wellid, and to further restrict the singular identifiers to be the wellids. A consequence of this would be that the definition of the singular identifier would be almost identical to the current definition of Unique Well Number.

Implementation

The OWI project is developing a version of the CWI database that attempts to implement the MNU identifier model in a stepwise fashion. Version 0 is simply a copy of the c4– tables served on the MGS's ftp server. Versions 1, 2, and 3 simply add the wellid and well locations. Version 4 introduces the o1– table series, which are to be mirrors of the c4– tables, beginning with o1id. Table o1id is just a copy of table c4id, with the addition of MNU and sMNU columns. Several temporary fields are also added to tables c4id and o1id so that non-compliant data and redundant records can be flagged and tracked. A temporary table, o1id_match is also created to document possible relationships. Version 4 also adds data integrity constraints on MNU identifiers, and only compliant data can be copied into table o1id.

In version 4.1, non-compliant data is modified so that it can be copied into o1id. The modifications are performed automatically, and are not guaranteed to be correct. The purpose is to generate an experimental data set that

prototypes examples of the more complicated MNU identifier relationships. Version 4.1 has to generate wellids for 'sets of wells', those wellids are *not* recognized outside of the OWI database as Minnesota Unique Well Numbers.

In a future version, the number of leading zeros used when storing W-numbers and H-numbers will be standardized in the identifiers table. Search tools will have to coerce search values to match the standardized format, but will not have to search all possible formats.

The final step in implementation would be education. Both maintainers and users of well data will need education about how to properly use MNU well identifiers. The most difficult lessons will most likely be to understand that the MNU model still provides functional uniqueness, and that wellids are not dependable identifiers outside of the CWI database. The most difficult lessons ought to be explaining sets of wells. How they are implemented will affect how easily they will be understood.

Conclusion

The MNU identifier model gives permanence to well identifiers making them suitable for use in sharing and in archives. It attempts to regularize several complex real-world identifier relations into its design so that standard data management concepts can be used to simplify handling them.

Functioning of the MNU identifier model depends on the presence of a public facing service to provide identifier relationships. That service could be provided by the Minnesota Well Index (MWI).

The MNU identifier model can be implemented in part or in full. If implemented only for the basic case that each wellid represents a single distinct well, then the MNU identifier model can provide permanence to MNU identifiers while being a barely noticeable departure from the Minnesota Unique Well Number identifier model.

The full MNU identifier model makes explicit the subtleties of two complicated identifier relationships, the one in which wellids can represent multiple wells, and the one in which they represent undetermined wells. This differs from the current practice which does not address such relationships directly or uniformly. Therefore the MNU identifier model is harder to learn at first, but later provides tools and guidance that should improve data maintenance and should enable more accurate data analyses.

The changes required from the current CWI c5– data table definitions are

kept to a minimum. Only the c5id table is modified, and the modification consists only of adding two fields: an integer field MNU that takes only a handful of integer values, and a boolean field sMNU that is either True or False. With additional changes to the c5– tables, the concepts in the MNU identifier model could be made even more explicit, but that is left as an exercise for the reader.

Implementing data uniqueness constraints for MNU identifiers is not difficult, but some of the existing data is non-compliant or ambiguous, and it will require manual data analysis to correct and classify that data. Three main tasks will be 1) to finish de-duplicating known duplicate wellids, 2) to correctly document redundant MNU identifiers currently documented in the remarks table, and 3) to correctly classify MNU identifiers and wellids that represent sets of wells.

Finally, both maintainers and users of well data will need education about how to use MNU well identifiers. Correct implementation and maintenance of identifiers is critical for correct and efficient interpretation of data.

Appendix A
Kinds of MNU identifier relationships

Several relationship diagrams are presented to show examples of real-world identifier relations. The examples use symbols for the identifiers, on the left and for the wellids, on the right. The symbols used are

C Minnesota Well and Boring Construction Record Numbers.

H Minnesota Well and Boring Sealing Record Numbers.

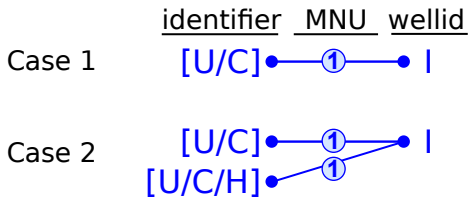
U Minnesota Unique Well Numbers, including historic identifiers.

I The wellid in CWI (Requires the existence of a well record in c5ix)

The symbols are given the suffix 'm', e.g. Cm or Hm, when they describe a set of wells. The symbols are given numeric suffixes to indicate different identifiers or wellid values. E.g. I1 and I2 represent two different wellids, and (U1, U1) and U2 represent three different well identifiers tied to I1 and I2.

Lines connecting identifiers to wellids represent rows in the o1id table. Each line is labeled with a circled number which is the integer type entered in the MNU column in o1id. (Sorry, please don't confuse the MNU identifier with the MNU type)

The choice of a singular MNU identifier for wells having multiple MNU identifiers is a secondary concern in these relationship examples, but you can imagine that the singular MNU identifier is the identifier connected to the wellid with a *horizontal* connecting line. Of course, in these diagrams a wellid can have only one horizontal connecting line.



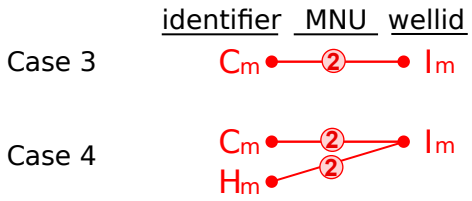
MNU identifier model cases 1 and 2, as described in the text.

Case 1 is a simple case of a well having only one MNU identifier. It is the only case among those presented here that explicitly allowed in CWI. [U/C] symbolizes that the identifier is either an assigned Minnesota Unique Well Number (U) or a Well and Boring Construction Record Number (C).

If the requirement for integer identifiers were kept, it could not be an H-number (H). The single connecting line indicates that there is exactly one row in o1id, and the circled number 1 indicates that the identifier is assigned MNU=1. Examples in CWI include identifiers 215010, 550010, 800010, and 1000010621.

Case 2 is a well with more than one MNU identifier. The types of identifiers in the first row are restricted to those that can be represented as integers, while the types in the second row are not. Each identifier is entered in o1id, so there are two connecting lines in the figure. Each appears exactly once in o1id with MNU=1. Examples in CWI include (208461 and 262010), (130522, 249227, and 258823), (13921 and H288420), (24253 and 24280), and (222051 and 803793). The latter two cases are interesting because they are among 87 wells that CWI maintains as dual rows in c5ix with cross references in c4id apparently because users have requested to maintain both identifiers. The last case represents a new Well Construction Record Number used on a well reconstruction report.

Cases 3 and 4 illustrate single Well Records that describe multiple wells, and single wellids assigned to represent the set of wells. In these cases there are no identifiers or wellids assigned to the individual wells.



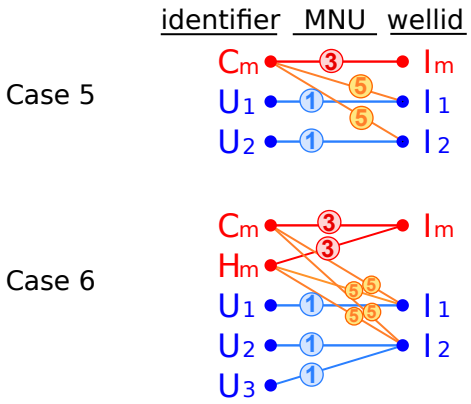
MNU identifier model cases 3 and 4, as described in the text.

Case 3 is the simplest case found in CWI at present. The 'm' in each symbol name (and the red color) indicates that the identifer or the wellid describe a *set* of wells. Examples in CWI include 720297 for 8 Dewatering wells, and 460530 for 4 Environmental wells.

Case 4 is like Case 3 except that there is also a Well and Boring Sealing Record for the set of wells. The fact that both identifiers link to the same wellid means that both records describe the *identical* set of wells. Examples in CWI include pairs (720297 and H229344) and (553830 and H63340)[1].

[1]The latter pair is interesting because the identifier H63340 is crossed out on the image

The uniqueness rules for MNU identifiers that are enforced in table o1id are enforced on any MNU identifiers labeled with MNU ∈ (1,2,3).



MNU identifier model cases 5 and 6, as described in the text.

Cases 5 and 6 are like cases 3 and 4, respectively, but now identifiers and wellids have also been assigned to represent the individual wells. In Case 5, identifier Cm and wellid Im describe the set of wells, and identifiers U1 and U2 and wellids I1 and I1 describe the individual wells. Note that although there are 3 wellids, there are only two actual wells. The relationships between set of wells and the individual wells in the set are provided by two new entries in the o1id labeled with MNU=4. This requires entering identifier Cm multiple times. That can be done without violating the MNU uniqueness constraint because the MNU field is assigned MNU=4. The 4 signifies a cross reference from a multi-well identifier to an individual wellid. [2] Identifiers having MNU=4 are subject to a different uniqueness requirement: That the combination (identifier, wellid) is unique, and that the identifier is also part of a relationship with MNU=2. Examples are known to exist but they are difficult to identify in the c4 tables.

Note that connections from U1 and U2 to Im exist, and could be entered, but they are not. Not entering them is an arbitrary choice to avoid redundant information and to keep the table entries as few and as easy to maintain and understand as possible. The connections from U1 and U2 to Im can be

stored in CWI, but is entered in table c5id. Crossing out *any* MNU identifiers on records should not be done.

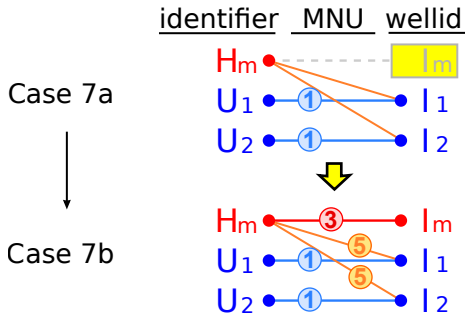
[2] An alternative approach that should be considered is to document such cross references in a new separate table.

reconstructed by following the links in order

$$\mathrm{U1} \xrightarrow{1} \mathrm{I1} \xrightarrow{4} \mathrm{Cm} \xrightarrow{2} \mathrm{Im}$$
$$\mathrm{U2} \xrightarrow{1} \mathrm{I2} \xrightarrow{4} \mathrm{Cm} \xrightarrow{2} \mathrm{Im}$$

.

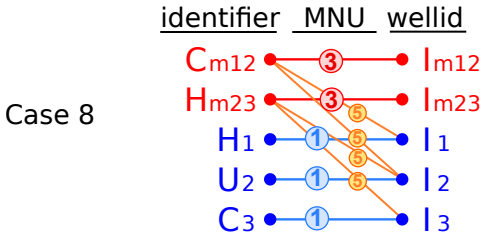
Case 6 is more complicated because there are more relations to show. But it presents nothing not previously described.



MNU identifier model case 7, as described in the text.

Case 7a illustrates a typical case of data entry in the current c5id table. The H-number from a Well and Boring Sealing Record describes a set of wells, but there is not a wellid created in c5ix to represent the set. The relations between Hm and the individual wellids are entered in c5id. This cannot be allowed in table o1id because then the identifier Hm would not be subject to the MNU uniqueness constraint. The needed wellid is shown in gray with yellow highlight. Case 7b shows that a unique definition is created by assigning a new wellid; it then resembles Case 5. Examples in CWI include (H72848 with 547905, 547906, and 547907) and (H307807 with 789134 and 764211)

Note in Case 7b that the singular identifier for wellid Im is still the H-number. If it is required that singular identifiers be integers, then another MNU identifier would have to be created. Since sealing records are common and in use by the well owners, it would make sense to define a distinct series of integer identifiers for the sealing records, perhaps computing them with a distinct first digit, and ending with the integer part of the H-number. E.g H307807 could be assigned 2000307807



MNU identifier model case 8, as described in the text.

Unfortunately things can get still more complicated when records describe intersecting but not identical sets of wells, as shown in Case 8. Fortunately, the above few rules handle these cases without further modification, and fortunately not many persons really need to understand how to enter or interpret these records. Also fortunately, the kinds of MNU identifiers allowed to describe sets of wells are presently limited to Well Construction Records and Well Sealing Records.

In Case 8, there is a site with 3 wells. Well Construction Record Cm12 describes the set of wellids I1 and I2, while wellid I3 is described by an individual Well Construction Record C3. The sealing of wellid I1 is described on an individual Well Sealing Record H1, and the sealing of wellis I2 and I3 is described by Well Sealing Record Hm23. In real-world cases the complexity can increase significantly with larger sets of wells and various intersections defined, but all relationship cases can be described as shown. The complexity is limited because wells are almost always limited to only one Well Construction Record and one Well Sealing Record. Examples in CWI of sets with partial overlap include well 720293 (16 wells) with H229346 (12 of 16).