# An Overview of the WHODrug C Format and Considerations for TMS

Sunil G. Singh
DBMS Consulting, Inc.
September 2005

# Acknowledgements and Introductions

- Many thanks to the OCUG for opportunity to present a tutorial related to the WHODrug C Format.
- My sincere thanks and appreciation to Daniel von Sydow of the World Health Organization, Uppsala Monitoring Centre for his invaluable contribution to this presentation.
- Many thanks to the audience members for attending.

# Agenda

- Overview of the C Format
  - Content of the C Format
  - Usefulness of the C Format
- Differences between B-2 and C Formats
  - Structural differences
  - Content differences between B-2 and C
- Supportability of the B-2 Format according to WHO-UMC
- Loading the C Format into TMS
- Migrating from the B-2 Format to the C Format in an existing TMS environment

# WHO Drug Dictionary History

- WHO Drug is a dictionary of known medicines maintained by the World Health Organization since 1968.
- It contains lists of all known manufactured drugs in every country that was ever reported to WHO.
- WHODrug identifies Generic Drugs (Preferred Terms) and non-Generic Drugs
- The dictionary also associates a drug with an Anatomical-Therapeutic Chemical (ATC) Classification; that is, the parts and systems of the human body where this drug might have an effect.
- The dictionary has changed structure (formats) three times in its history, the most recent in 2002 with the introduction of the C Format, which associates every Drug to an ATC code.

# WHO's Three Dictionaries

- The WHO Drug Dictionary, WHO Drug Dictionary Enhanced, and WHO Herbal Dictionary are different products; the difference between them are the content.
  - WHO-HD contains herbal products only
  - WHO-DD is the same WHODrug dictionary which has existed previously
  - WHO-DDE contains the same types of products as the WHO-DD but with the addition of a large number of new drugs from IMS Health.
- All three dictionaries are provided in the three different FORMATS - C, B-1 and B-2. Therefore loading considerations for WHODD are also valid for WHODDE and WHOHD.
- There are a few minor differences in the use of a few fields between WHODD and WHOHD.

# WHOHD Content

- The WHO Herbal Dictionary contains all products that only include ingredients of natural origin.
- Products that contain a combination of conventional substances and herbals will be included in the WHO Drug Dictionary and the WHO Drug Dictionary Enhanced.
- All entries in the WHO Herbal Dictionary are coded with the Herbal ATC classification.

# WHODD and WHODDE Content

- The WHO Drug Dictionary stays the same, herbals that are already included will stay there. The herbals in WHO DD will not be classified with the Herbal ATC classification.
- The WHO Drug Dictionary Enhanced contains all entries in the WHO Drug Dictionary plus a large number of entries from IMS Health. The reason why it is a separate product is that customers need to sign a specific agreement to be allowed to access the new data.

# Overview of the C Format

- The C Format dictionary is a dictionary of Medicinal Products

- Medicinal Product
  - A unique combination of
    - Name
    - Name Specifier
    - Market Authorisation Holder
    - Country
    - Substance and Strength (Ingredients and Units)*
    - Dosage form*
    - Drug Code

* NOT in the Medicinal Product Table!

# Content of the C Format

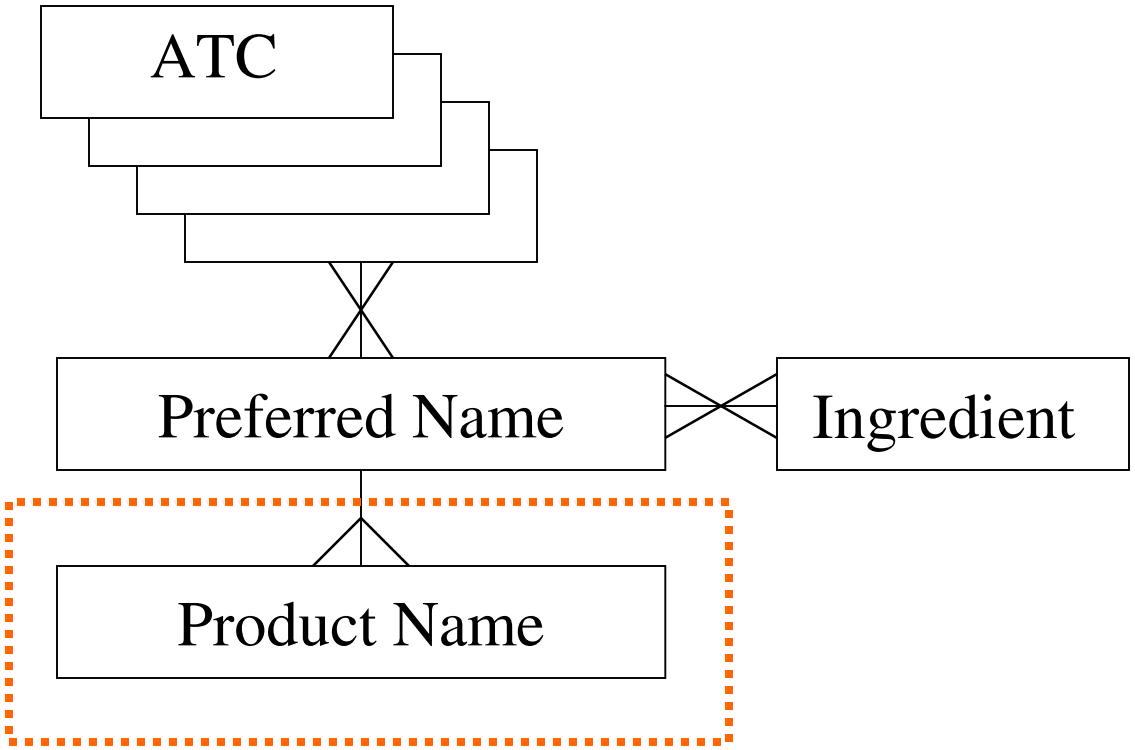
- The WHODrug dictionary contains many levels including:
  - Drug Names
  - Anatomical Therapeutic Chemical (ATCs) Classifications
  - Ingredients
  - Source Codes
  - Manufacturers
- Total number of entries in WHO Drug Dictionary Enhanced this quarter:

| | | |
|---|---|---|
| The C format: | 266 | 109 |
| The B-1 format: | 184 | 141 |
| The B-2 format: | 80 | 107 |

- Total number of entries in WHO Drug Dictionary this quarter:

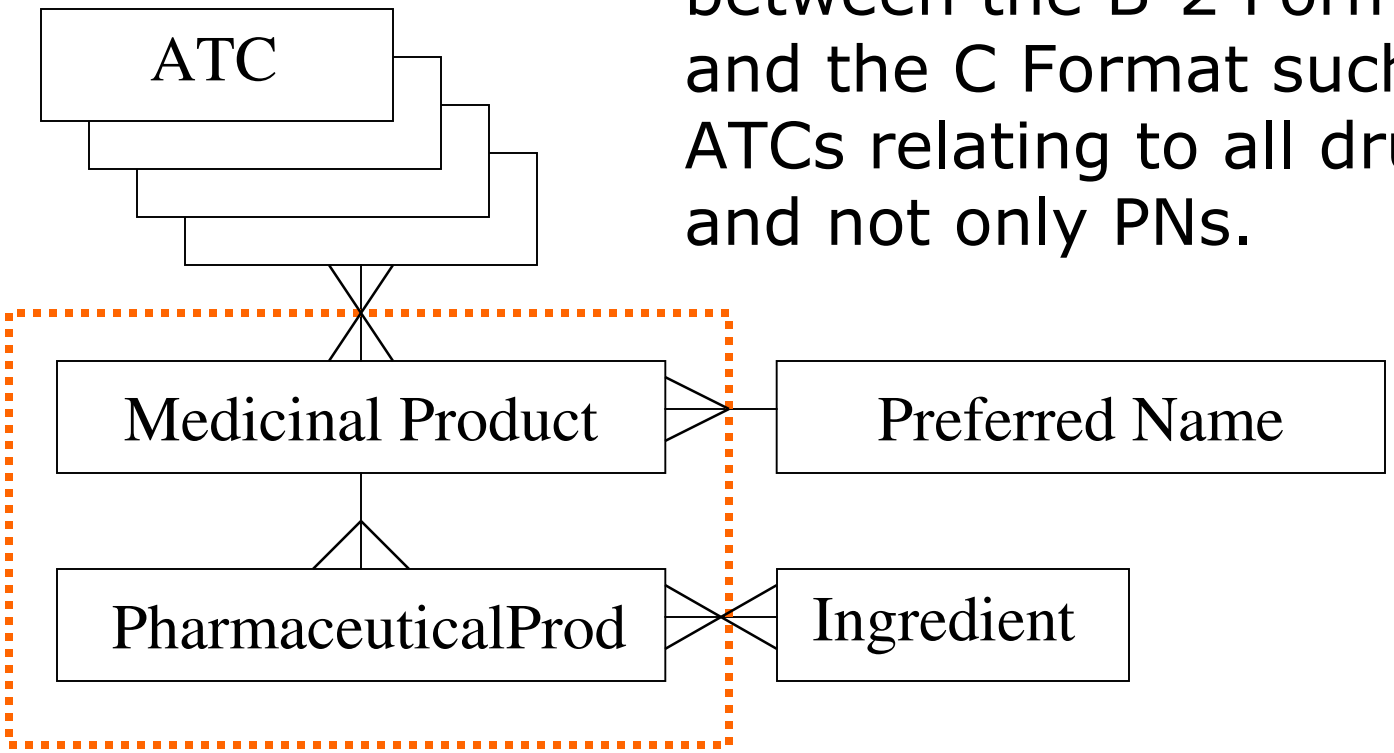
| | | |
|---|---|---|
| The C format: | 134 | 207 |
| The B-1 format: | 114 | 654 |
| The B-2 format: | 58 | 859 |

# WHODrug Dictionary the B-2 Format



# WHODrug Dictionary the C Format

Structural differences between the B-2 Format and the C Format such as ATCs relating to all drugs and not only PNs.



# Content Differences Between B-2 and C: Name

WHO Drug Dictionary B-2 Format

- Distributed for over 20 years
  - It is a dictionary of drug names, where a name can be searched and translated to coded information.
  - It consists of mainly active ingredients, drug codes (which represents active ingredients and salts/esters) and Anatomical Therapeutic Chemical Classification.
- The drug name appears only once
  - A drug name is added the dictionary at the first occurrence of the name.
- **Please Note:** The B-2 Format was made completely country independent in the March 1, version 2005.

# Content Differences Between B-2 and C: Country

WHO Drug Dictionary C Format

- The C Format allows for country specific information
  - It is possible to see which drug names appear in a specific country.
  - This information is especially relevant for certain types of products; where the same product names are marketed in different countries with different sets of ingredients.
  - In the B-2 Format the coder will not be able to determine which version of the drug is used in a certain country, but this information is available in the C Format.

# Content Differences Between B-2 and C: Dosage Form and Strength

The C Format contains more information than the B-2 Format; dosage form and strength. The UMC has put more focus on populating the dosage form information than the strength information for two reasons:

- The dosage form information is relevant to the analysis of clinical data.
  - Types of reaction may vary depending on the type of administration; local versus systemic effects, and there could be different types of reactions to a sustained release tablet compared to a regular tablet.
  - Inadequate dosage forms may explain adverse reactions; Esophagus Ulcer caused by capsules that weren't swallowed properly.
- Sometimes the same trade name is available in different dosage forms, with different ingredients.
  - The suppository could contain additional ingredients, or different salts of the substance.

# Content Differences Between B-2 and C: Drug Code

In the B-2 Format, the Drug Code,unique system code, describes the active ingredient(s), the salt/ester and the product name.

- The code is very useful for analysis, but it causes the following problems for data management:
  - The Drug code is affected when a product formulation is changed; one of the active ingredients is replaced by another, or a different salt of a substance is used.
  - The Drug code is affected when corrections are made; if a drug has been included in the dictionary with an incorrect salt or substance and later corrected.
  - The Drug code is affected when the name changes for various reasons. This means that the system has neither a code nor a text that is completely stable. (although these changes are exceptions and are not very common).

# Content Differences Between B-2 and C: Medicinal Product ID

WHO Drug Dictionary C Format

- Introduces the Medicinal Product ID.
- The reasoning behind introducing this key is that this logical, unique code doesn't have an intrinsic meaning and it will remain stable over time.

# Content Differences Between B-2 and C: ATC Coding with B-2

- Both the B-2 Format and the C Format contain ATC classification.
- WHO Drug Dictionary B-2 Format
  - All products are coded with the same ATC codes as its preferred name (an active ingredient or unique combination of active ingredients).
  - For example, all products containing Acetyl Salicylic acid will be coded with the following ATC codes:

| | |
|---|---|
| A01AD | LOCAL ORAL TREATMENT |
| B01AC | PLATELET AGGREGATION |
| N02BA | ANALGESICS AND ANTIPYRETICS |

# Content Differences Between B-2 and C: ATC Coding with C

WHO Drug Dictionary C Format

- A specific product is coded with the ATC code that reflects the most common use of the product.
- For example, an Acetyl Salicylic acid product used mainly as a painkiller would be coded with the N02BA code.

# Supportability of the B-2 Format according to WHO-UMC

- The UMC stated that the WHO Drug Dictionary B Formats will be discontinued in the end of 2005.
  - This date was set based on input from a customer survey.
- Many customers notified UMC about the difficulties of moving to the C Format, which involves the modification of software and the development of new Standard Operating Procedures for coding.
  - Many customers cannot make these modifications before the end of 2005.
- Simultaneously, new initiatives were put forward in the area of drug coding
  - New WHO Herbal Dictionary.
  - WHO Drug Dictionary Enhanced.
  - Possible future standardisations by the ICH.

# Supportability of the B-2 Format according to WHO-UMC (2)

**In the light of these developments, the UMC decided not to discontinue the B Formats.**

- The B Formats will be produced together with the C Format for the foreseeable future.
- **Please Note:** The new WHO Herbal Dictionary and the WHO Drug Dictionary Enhanced will be produced in both B and C Formats.

# Loading the C Format into TMS

- Explain challenges to loading the WHO Drug the C Format in TMS 4.5.
- Identify the key decision points that must be addressed before loading.
- Provide suggestions for possible loading and configuration options.
- Provide suggestions on possible future enhancements and changes both to the WHODrug C Format and TMS.

# Differences in the WHODrug C Format Affecting TMS

- The Drug names themselves are **not** unique in the C Format.
- ATC codes are now associated to every Drug Name in the C Format.
- A Pharmaceutical Product level, which contains the Pharmaceutical Form (PF), was introduced in the C Format.
- All ingredients and their amounts were introduced in the C Format.
- The Medicinal Product ID (MP ID), which represents 7 drug attributes, now uniquely identifies a drug
  - (Drug Name, Name Specifier, Country, Manufacturer, All Ingredients w/ Strengths and Units, Pharmaceutical Form Drug Code (DrgRecNum+Seq1+Seq2)).

# Loading Considerations

- Since the drug name is not unique in the C Format, the drug name alone can not be loaded as the Classification level in TMS. Therefore, the drug names must be made unique somehow.
- In making drug names unique in the C Format, the TMS built-in automatic matching would potentially be diminished. Some considerations have to be made for preserving TMS auto encoder efficiency.
  - There should be an entry with only the Name as he Classification Term and Drug Code as the DICT_CONTENT_CODE.
  - Sometimes there are two different Drug Codes (sets of ingredients) for the same Name. In these cases, the TMS coder needs to view the higher levels of the dictionary to find the difference between the entries - it could be country or pharmaceutical form.

# Loading Considerations (2)

- Considerations to preserve TMS auto encoder efficiency
  - In the March 1 2005 version of the in B-2 Format, the /.../ was added to all names that appeared with more than one drug code including Preferred name entries XXXXXX01001. Approximately 14% of the names needed the additional /.../ code in order to make them unique. The reason why the /.../ code is added is that there is AT LEAST one more entry with the same name but different drug codes. That means that at most 7% of the names are "non-unique".
  - In the June 2005 version of the in B-2 Format, the preferred name entries are left without the /.../ code in order to make autoencoding possible.
  - A similar solution for C users of TMS may be possible in the future.

# Loading Choices

- Option 1: Use the Medicinal Product ID itself to make the Drug Names unique in the classification level.
- Option 2: Use the logical expansion of the Medicinal Product ID to make the Drug Names unique in the classification level and possibly populate a VTA Level with Drug Names only.
- Option 3: Add an additional level to store the Drug Names only as part of a Classification Group in the TMS WHODrug structure.

# Option 1 : Medicinal Product ID at Classification Level

Advantages

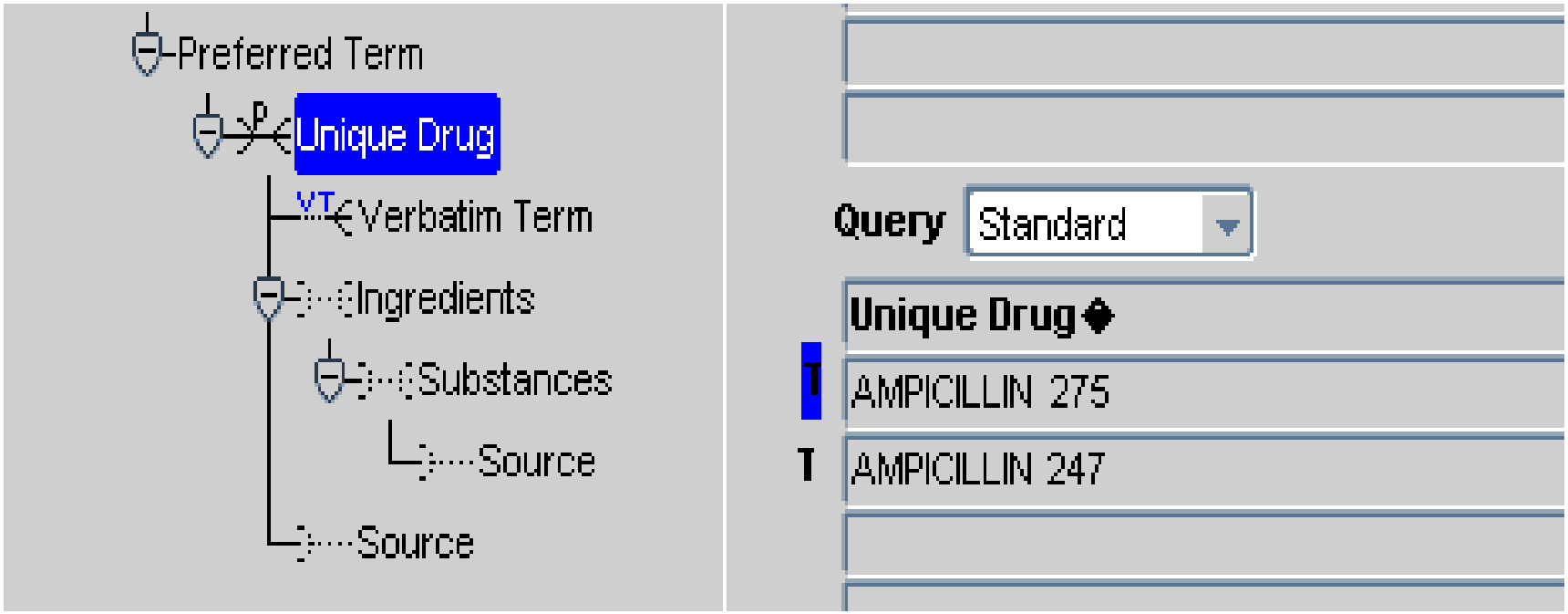
- Easy to load.

Disadvantages

- Auto encoding would not be possible.
- Coders would not have information needed to select correct VTA.

- Another suggestion is to add the MP_ID to only non-unique drug terms. However, this still leaves many terms (10,000+) which will not auto encode, and therefore, are less likely to be used.

# Option 1 : How it Looks While Coding




DrugName MP_ID

# Option 2 : Use the Logical Expansion of the MP ID

Advantages

- Information is available for coders to select appropriate VTA.

Disadvantages

- Nothing auto encodes.
- Load script is more complicated and takes longer.
- Field length may require > CHAR 300.

# Option 2 : Auto Encoding Implications

- Load Verbatim Term Assignments (VTAs).
- This also allows coders to use the filter buttons in TMS Omission Management to choose the VTA Level and only code on the Drug Names if desired.
- Problem - Over 10,000 drug names are not unique.
- Do you have to manually code all 10,000+?
  - Yes and No !

# Option 2 : Manually Code All Duplicate Drug Names

## Advantages

- Control of the codes – you can select certain drugs from specific countries, or manufacturers, or ingredients, etc.
- You could reload same VTAs, once they are selected to new versions of the dictionary.

## Disadvantages

- As each version is released, you will need to repeat this exercise.
- How long will it take your team to code 10,000+ terms?
- Some of these terms you will never see in a study, but you will spend a lot of time on them initially.

# Option 2 : Or, Don't do it Manually

- Load only the Unique Drug Names as verbatim terms.
- Code the others as they are encountered as verbatim terms.

## Advantages

- Over 40,000 will be able to have VTAs loaded.
- You only spend time on those you need.

## Disadvantages

- You may give up consistency in decision making if this is done over time.
- Many of the most common drugs encountered are in this group.
- You need to repeat this with each new version of the dictionary.

# Option 2 : Or, Do it Systematically

Advantages

- Same script can be used for each new version of the dictionary.
- Logic can be applied that is consistent across all term choices.
- The script will run faster than your team can do the work!

# Option 2 : Or, Do it Systematically (cont)

Disadvantages

- Decisions still need to be made on the logic to be used.
- Some terms will not have VTAs because the same drug name by different. countries/manufacturers are really different drugs.
- TMS loading Script development is complex and will take some time to run!
- Additional code must be added into the TMS loading script to take into account PF and strength.

# Which Drugs Should Have VTAs?

- Drugs having the same DrgRecNum and Seq1 and can have a VTA selected.
- The same DrgRecNum and Seq 1 mean the drug is the same drug with the same Preferred term and the same ingredients.

- **Please Note:** WHODrug will continue to support the DrgRecNum and Seq numbers (see the document titled The New C Format: New Features that accompanies each version of the dictionary).

# Reason for Multiple Drug Record Numbers

- A strategic decision by a manufacturer to change the active ingredients to improve the product over time, but keep the same Drug Name due to market share and brand recognition.

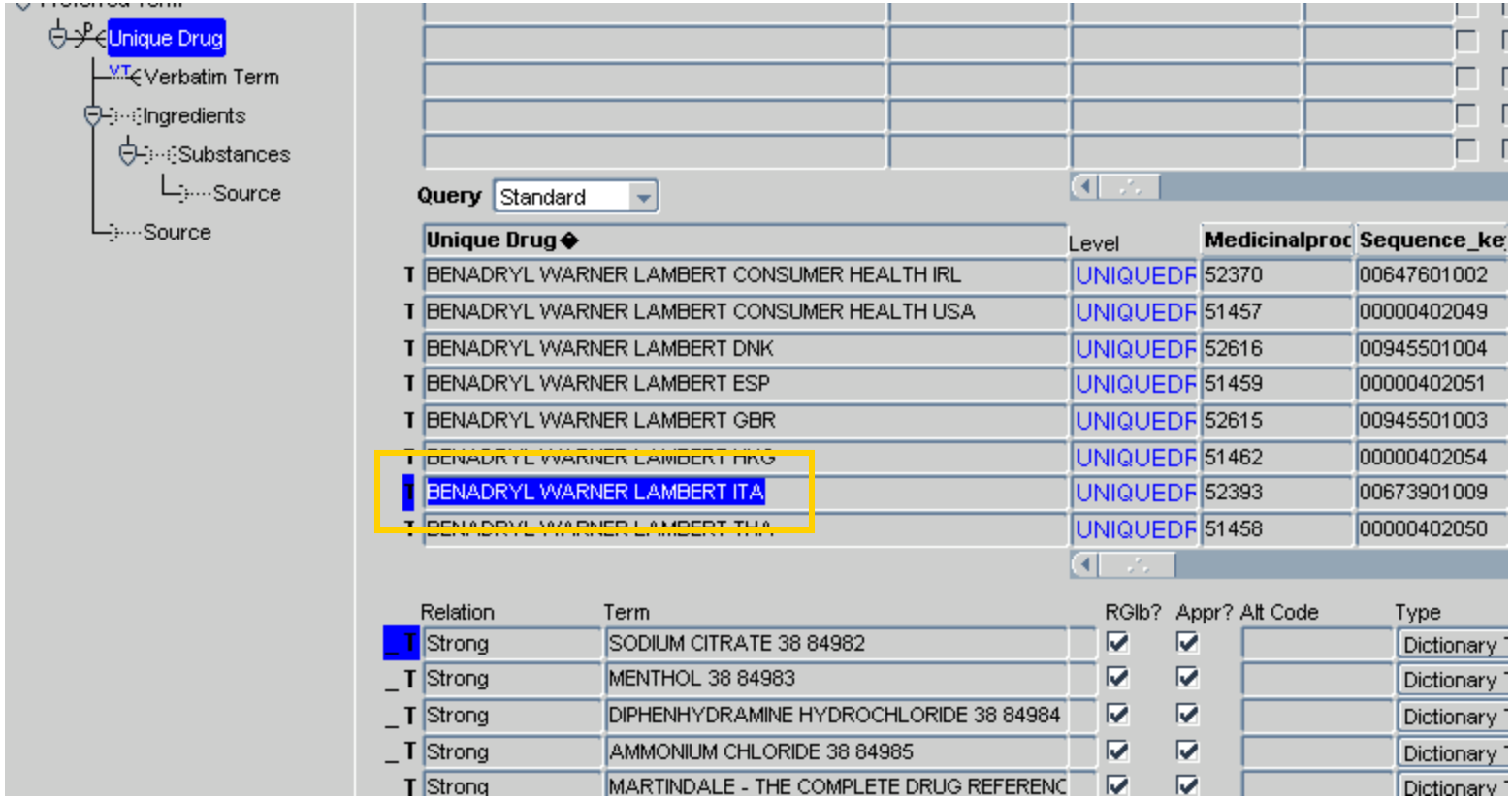
- The lack of availability of some active ingredients in some countries or geographies, including cases where the raw materials are not available or are banned by a country for human use or import.

# Reason for Multiple Drug Record Numbers (2)

- The purchase or acquisition of one company or brand by another combined with a strategic decision to keep the same brand recognition and market share purposes, but to also change or improve the drug which might change the active ingredients.
- The lack of enforcement of intellectual property rights or patents in some countries, where the same Drug Name is used illegally and manufactured with completely different ingredients. WHO-UMC is still obligated to report the creation and use of these drugs.

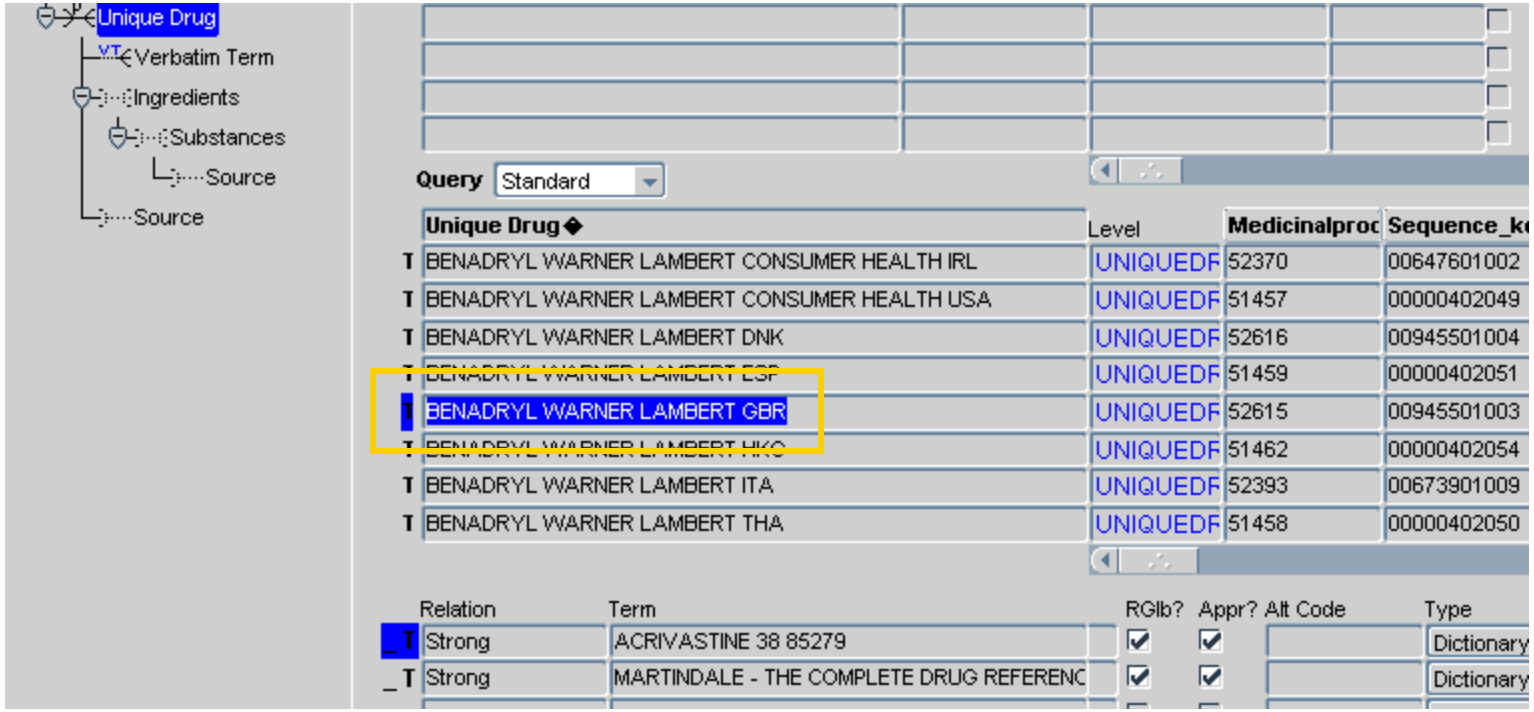
# Same Drug Name in Different Countries

## Benadryl in Italy



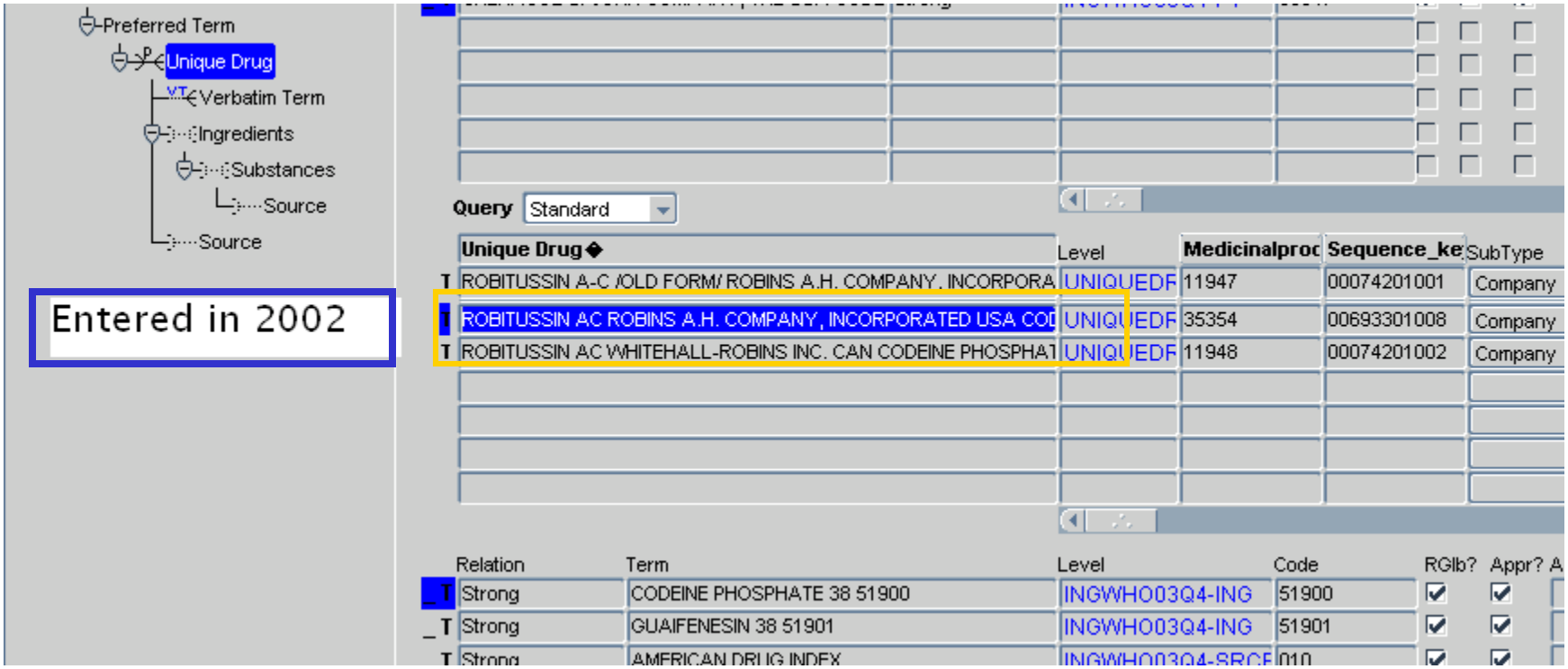
# Same Drug Name in Different Countries (2)

## Benadryl in the United Kingdom



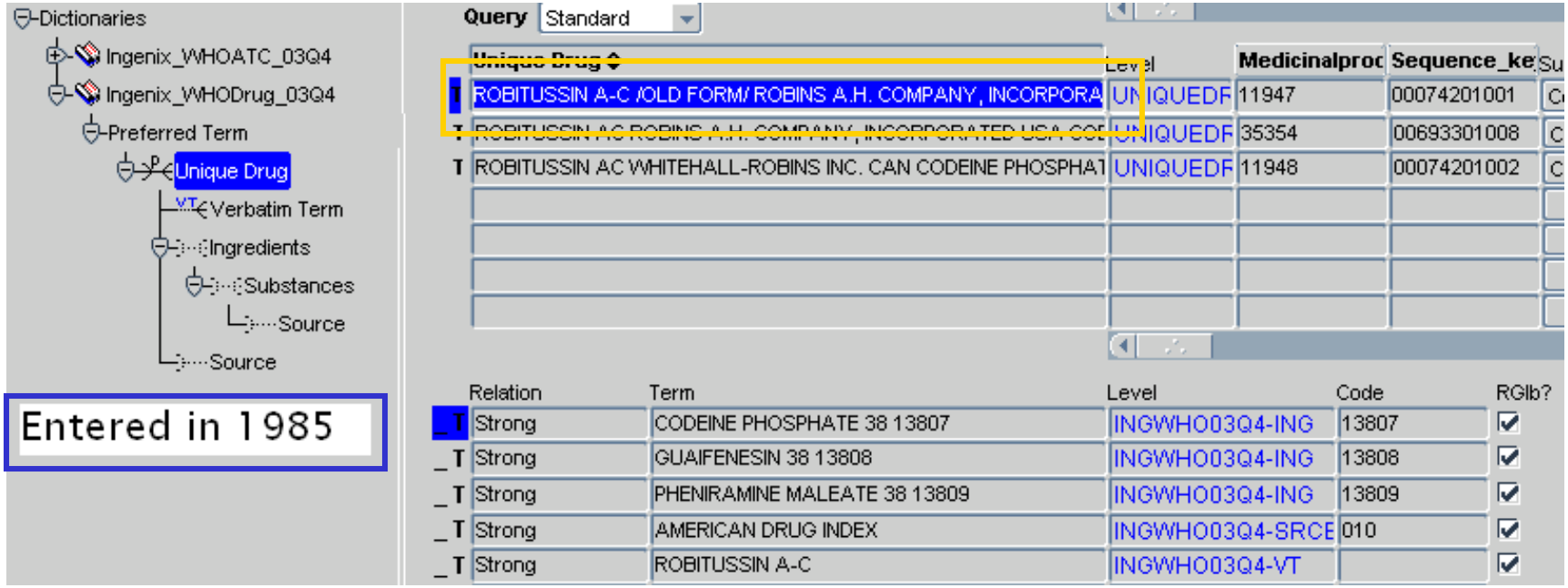
# Changes in the Drug

## Robitussin AC

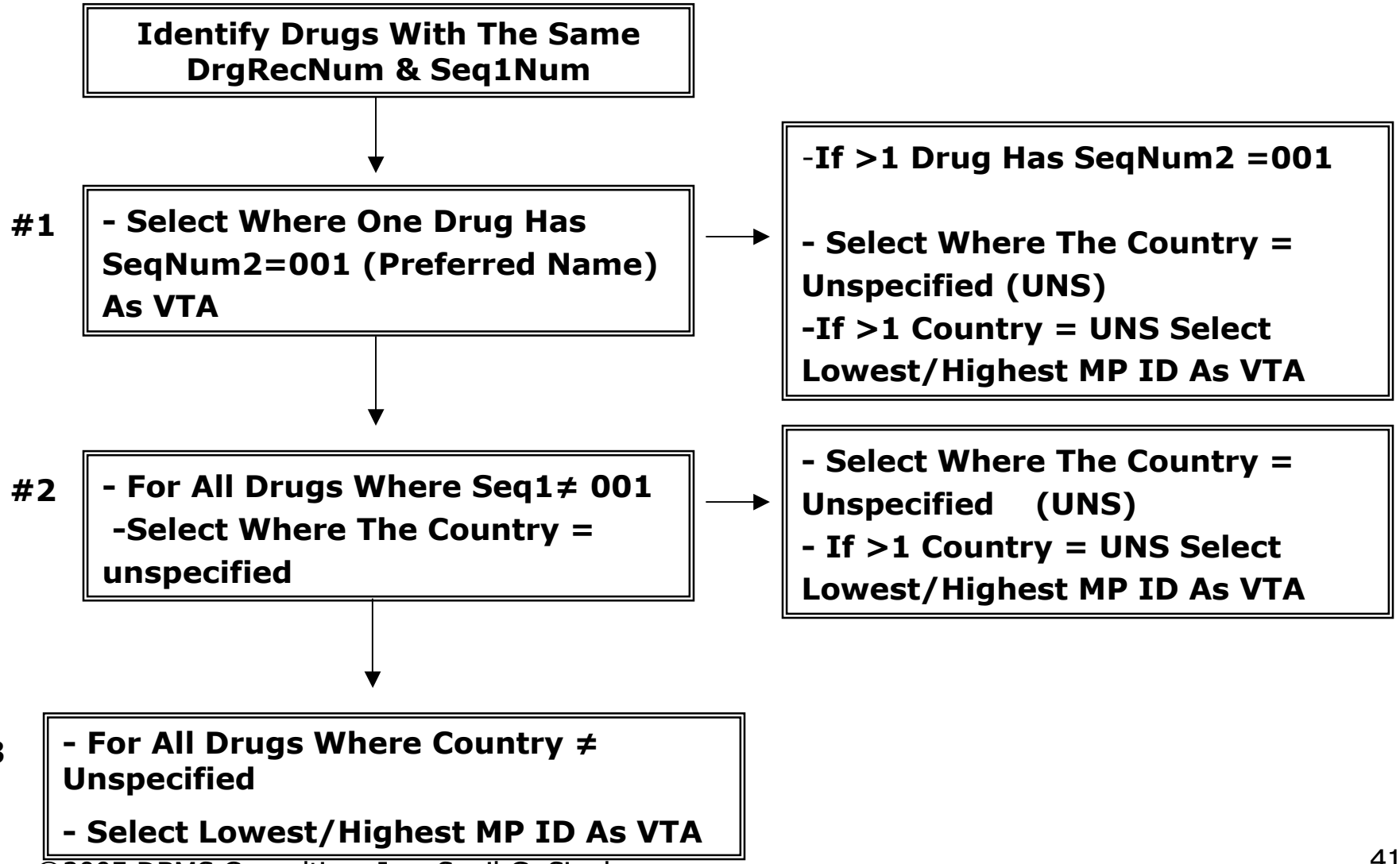


# Changes in the Drug (2)

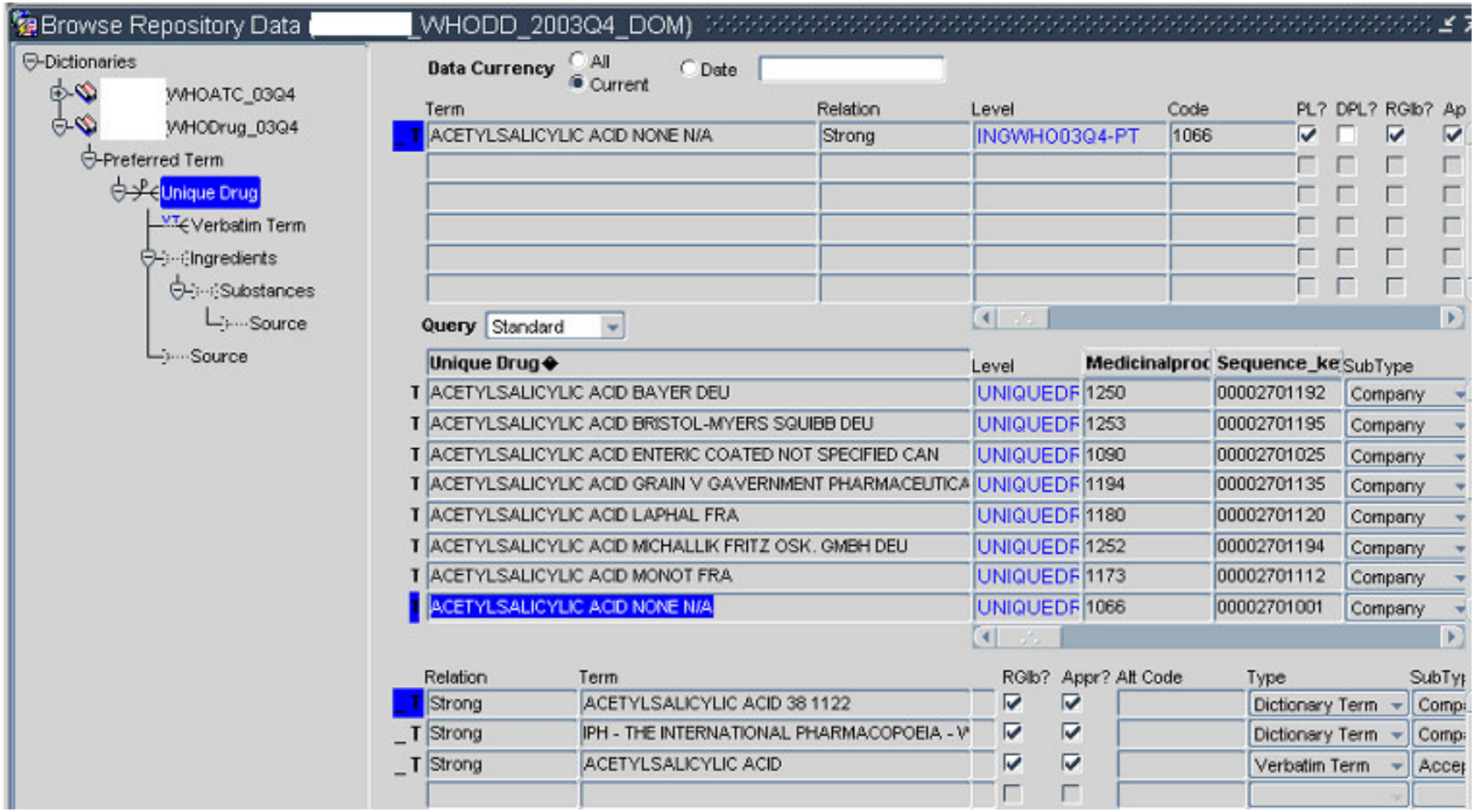
## Robitussin AC



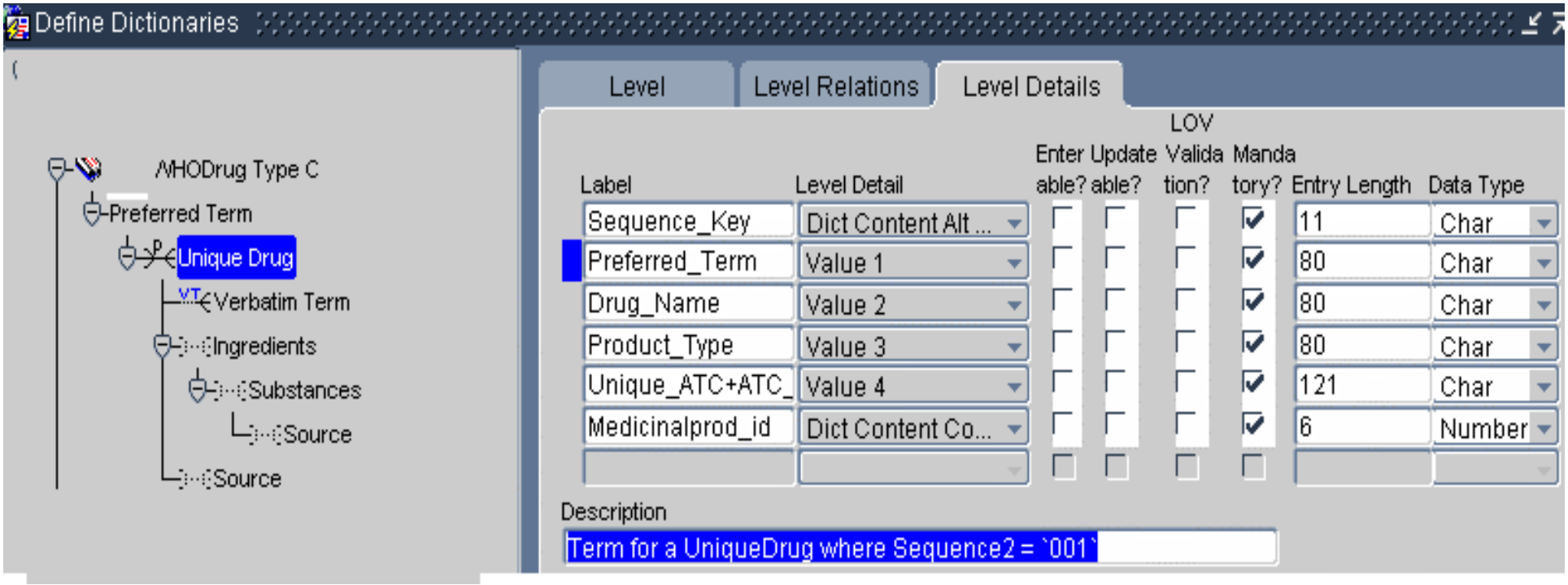
# A Suggested Decision Tree for Selecting VTA



# Classification Level in WHODrug C Format



# Derivable Level Details with Drug Name and Preferred Drug Name in WHODrug C Format



# Option 3 : New Level in TMS

- Load Classification level with Drugname + MPID.
  - This level is now unique and since your are not going to try to "code" anything to it, it does not need additional description.
- For all drugs at the classification level, determine if there are multiple drug names and, for those, if they have the same or different DrgRecNum and SeqN 1.
- If the Drug Name is unique, load it at the middle level intermediate.

# Option 3 : New Level in TMS (2)

- If the Drug Name is not Unique, but all the DrgRecNum|SeqN1's are the same, verify that the PF, Strength and Ingredients are all the same and then choose one to DrugName_MPID to represent the rest and load that at the middle intermediate level.
- You can use the logic presented earlier or some other similar, reproducible logic for selecting which one will represent all these drugs that are the same.
- This would be a stored relation in the TMS dictionary structure.

# Option 3 : New Level in TMS (3)

- Additionally, some other extensions of this option might make it more useful:
  - Where an entry is made the Intermediate Middle Level, also make a corresponding entry to a VTA level (combining aspects of Option 2) –OR-
  - Create a Classification Group which contains the Intermediary Level plus the level with Drug Name + MP_ID
  - All other non-unique drug names including multiple cases of the same drug name with different DrgRecNums| Seq1 can be loaded to the new mid-level with Drugname_PharmaceuticalForm_Strength_ Ingredients.
  - This level will now give you the information your coders need to make decisions on the correct one to chose for coding.
  - A primary link could be considered between the Unique Drug classification level and the Intermediate Level.

# Option 3 : Other Considerations of Loading New Level

- Loading an additional dictionary level would require structural changes to the existing relations between the Unique Drug Names classification level and the intermediate Middle Level in order to change the default Drug Name/MP_ID, that is, the default coding behavior. These changes would be made through Maintain Repository Data, which can lead to other data-related mistakes or exposures of bugs which can be very difficult to repair.
- Some thinking from a validation perspective might be that the structure and data of a vendor dictionary should never be modified. In this context, modification of VTAs is not the same as modification of the WHO-UMC's dictionary data or dictionary structure.

# Dictionary Updates and Reducing Data Scope

- One consideration is whether or not all of the Drug data should be loaded. Why not parse all of the Drug Names only and simple load these Drug Names?
  - Not Loading the MP_ID or loss of the MP_ID will make updating this dictionary very difficult. This is because the default TMS APIs for updating the dictionary, TMS_LOAD_DICTIONARY.MigrateRelations and TMS_LOAD_DICTIONARY.MigrateTerms expect a unique DICT_CONTENT_CODE in the dictionary which comes from the vendor which can be compared with queries against the vendor source data to determine what DICT_CONTENT_CODEs to insert/update/delete.
  - Additionally, during the dictionary load process, it is not required to specify a DICT_CONTENT_CODE nor is uniqueness enforced! But during update calls using the TMS_LOAD_DICTIONARY API, it is a de facto expectation.

# Dictionary Updates and Reducing Data Scope (2)

- This means not having the MP_ID for all of the WHODrug source data will make updating very difficult. Calls to TMS_USER_MT_DICTIONARY for updating, inserting and deleting terms will have to be made on a separate basis, without the benefit of the TMS migration APIs.
- Additionally, if only part of the drug data is loaded (a reduction in the data scope), it may be possible to make a validation argument that the dictionary loaded in TMS was not actually a representation of the WHO-UMCs WHODrug dictionary, but a customized dictionary which is a proprietary to a single organization, which may introduce some additional validation requirements.

# Suggested TMS Enhancements to Facilitate Loading WHODrug C

- Not Requiring Classification Level to be Unique, but instead requiring the combination of the TERM_UPPER along with LEVEL_1 to be unique?
- Allowing configurable column lengths in TMS_DICT_CONTENTS for TERM and TERM_UPPER, so that customers which require a longer Term field (such as Option 2) can extend their classification content past 300 Characters?
- Create a TMS codelist which is default enables which requires dictionaries to be loaded with DICT_CONTENT_CODEs which are not null and unique?

# Migrating from the B-2 Format to the C Format

## Tools for Migration

- Mapping of datafields Old to New, New to Old.
- Mapping of key descriptors.
- Documents describing the structures, code systems, new features etc.

## Tools for Quarterly Versioning

- List of changed Drug Code (drug record number, sequence 1 and sequence 2). E.g.. when a product is coded under new salt - Ampicillin changed to Ampicillin Sodium.
- Medicinal Product ID listed together with old and new drug code.

# Migrating from the B-2 Format to the C Format (2)

Tools for Quarterly Versioning:

- List of moved Medicinal Product IDs. Used when a product is coded under new salt and an entry already existed.
- Old and new Medicinal Product ID listed together with old and new drug code.
- List of changed drug names.
- Drug code listed together with old and new drug name.
- Changed ATC codes, and Deleted ATC codes.
- Drug Code, ATC code and year/quarter.
- Yearly ATC revision.

# Migrating from the B-2 Format to the C Format (3)

Tools for Quarterly Versioning:

- Changed CAS numbers, and old CAS numbers.
- Drug Code, old and new CAS numbers.
- Need for a table with Drug Code and the corresponding Medicinal Product ID?
  - An "additional column" in B-2.

# Mapping Process in TMS

- Since the structure of an existing TMS dictionary with coded data can not be changed, a mapping must be established to associate the existing TMS Dictionary to the new the C Format data.
- This involves:
  - Updating the Term definition in the Classification Level.
  - Mapping Drug_Code to the Medicinalprod_id.
  - Creating new VTAs since the Terms are no longer Unique.
  - Updating DICT_CONTENT_CODE keys where they have changed.
- Even though the relation between ATCs is now at the Medicinal Product level, still relations of ATCs to Preferred Names can be established since Drug Recnr, Sequence1 and Sequence2 are provided.

# Changes to Key Mappings

- The actual keys change for the Source, Organization, and Country.
- Translation tables are supplied to update the keys, which can update the existing TMS Structure with tms.tms_user_mt_dictionary.UpdateContent.
- These Keys were generally used in the existing TMS WHODrug B-2 format structure:
  - MAN Level: Dict_content_code Company_code becomes Organization_id.
  - MAN Level: Value_1 Country becomes New Country Code.
  - DDSCR Level: Dict_content_code Old Source_Code becomes New Source_code.
- Dict Content Code is essential for updating relationships for these levels and must be corrected.

# Mapping the B-2 Format to the C Format

| Sources of drugs (DDSRCE) | Source (SRCE) |
|---|---|
| Source code | SRCE.Source code (Translation table supplied) |
| Source | SRCE.Source |
| Country code | SRCE.Country code |
| **Manufacturers (MAN)** | **Organization (ORG)** |
| Company code | ORG.Organization_Id (Translation table supplied) |
| Name | ORG.Name |
| Country code | ORG.Country code |
| **Country codes (CCODE)** | **Country code (CCODE)** |
| Country code | CCODE.Country code (Translation table supplied) |
| Country name | CCODE.Country name |

# Mapping the B-2 Format to the C Format (2)

| Drug dictionary (DD) | Medicinal + Pharmaceutic product (MP + PP) |
|---|---|
| Drug record number | MP.Drug record number |
| Sequence number 1 | MP.Sequence number 1 |
| Sequence number 2 | MP.Sequence number 2 |
| Check digit | NA |
| Designation | NA |
| Source year | MP.Source year |
| Source code | MP.Source code |
| Company code | MP.MA Holder |
| Number of ingredients | PP.Number of ingredients |
| Salt/ester code | NA |
| Year, quarter | MP.Create date |
| Drug name | MP.Drug name + Name specifier |

| ATC classification (DDA) | Therapeutic group (THG) |
|---|---|
| Drug record number | NA |
| Sequence number 1 | NA |
| Sequence number 2 | NA |
| Check digit | NA |
| ATC code | THG.ATC code |
| Year, quarter | THG.Create date |
| Official ATC code | THG.Official ATC code |

| Ingredients (ING) | Ingredient (ING) |
|---|---|
| Drug record number | NA |
| Sequence number 1 | NA |
| Sequence number 2 | NA |
| Check digit | NA |
| CAS number | SUN.CAS number |

# Questions / Contacts

**Sunil G. Singh**
DBMS Consulting
e-mail: singh@clinicalserver.com
Call: US 001-860-983-5848

- Electronic copies will be posted on the OCUG Intranets Site and www.clinicalserver.com

**Daniel von-Sydow**
World Health Organization, Uppsala Monitoring Centre
e-mail: daniel.von-sydow@umc-products.com

**World Health Organization, Uppsala Monitoring Centre**

Web address: http://www.who-umc.org