

The WHO Drug Dictionary Types and Formats and Loading Considerations in TMS

Daniel von Sydow

Anders Hansson

Sunil G. Singh



Acknowledgements and Introductions

- Many thanks to the OCUG for opportunity to present a tutorial related to the WHODrug Dictionary Types and Formats.
- Many thanks to Anders Hansson and Daniel von Sydow of the World Health Organization, Uppsala Monitoring Centre.
- Many thanks to the audience members for attending.



Agenda

- Part I: Overview, Content and Usefulness of the WHODrug Dictionary Types
 - WHODD
 - WHODDE
 - WHOHD
 - Combined Files of WHODDE and WHOHD
- Part II: Overview of the WHODrug Formats
 - B1/B2 Format and C Format
 - Uses of C Format
 - Differences between B2 and C Format
- Part III: Loading and Updating the B2 and C Formats into TMS
 - The ATC Derivation Problem and Options in TMS
 - C Format Loading Considerations in TMS
 - Update Considerations for B2 Drug Code



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Part I: Overview, Content and Usefulness of the WHODrug Dictionaries



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 provides a unique MP_ID and associates EVERY Drug to an ATC code



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



WHODrug Dictionary History (2)

- Until 2002 there was only one format
- Until 2005 there was only one type
- Historical data is often coded with
 - dictionary type: WHO Drug Dictionary
 - dictionary format: B-2



WHODrug Dictionary Types

- 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.
 - WHO-Combined contains the content of WHO-DDE and WHO-HD without overlaps in data.
- 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.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



WHO Drug Dictionary

- The WHO Drug Dictionary contains medicinal data that has been reported from National Centers
- In order to populate the dictionary with all products in all countries the UMC entered a collaboration with IMS Health
- Increased the number of names with 300% (B-2 entries)



WHO Drug Dictionary Enhanced

- Collaboration with IMS required a new agreement with the subscribers
- WHO Drug Dictionary Enhanced was produced as a separate dictionary type
- Subscribers that have not upgraded can still use WHO Drug Dictionary – without the IMS data
- New customers get WHO Drug Dictionary Enhanced



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



WHO DDE - Uses

- More names – increased chance of finding a ‘direct hit’.
 - Less manual work
- Reduced need for taking chances and “googleing” – higher quality of data.
- Non-unique names may have “siblings” only in WHO Drug Dictionary Enhanced



WHO DDE - maintenance

- The WHO DDE grew dramatically during 2005-6.
- It continues to grow with data from IMS – new launches and new formulations
- Modified formulations are also reported from IMS
- Other sources of data are also entered into WHO DDE



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.



WHO Herbal Dictionary

- A need for special classification of herbal products – botanic instead of chemical.
- The Drug Code identifies plants and parts of plants instead of molecules and salts
- ‘CAS number’ (substance ID) identifies plants etc
- Herbal ATC contains additional groups



WHO Herbal Dictionary - Uses

The 'chemical environment' contains also the herbal remedies the patients take.

Trade names for herbal products can be found.



Combined dictionaries

- WHO Herbal Dictionary is distributed seamlessly integrated with WHO Drug Dictionary and WHO Drug Dictionary Enhanced
- All files contain a mix of herbals and conventional products
- ATC files contain a mix of ATC and HATC
- No overlaps!



Part II: Overview of the WHODrug Formats

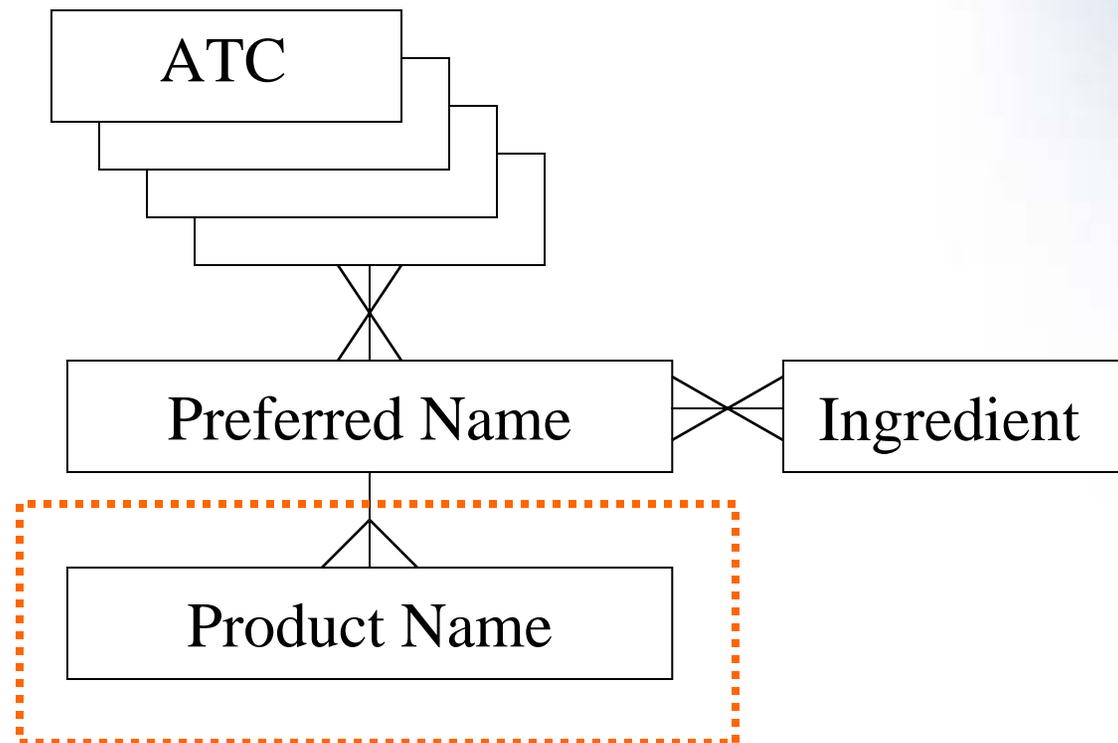


Dictionary Formats

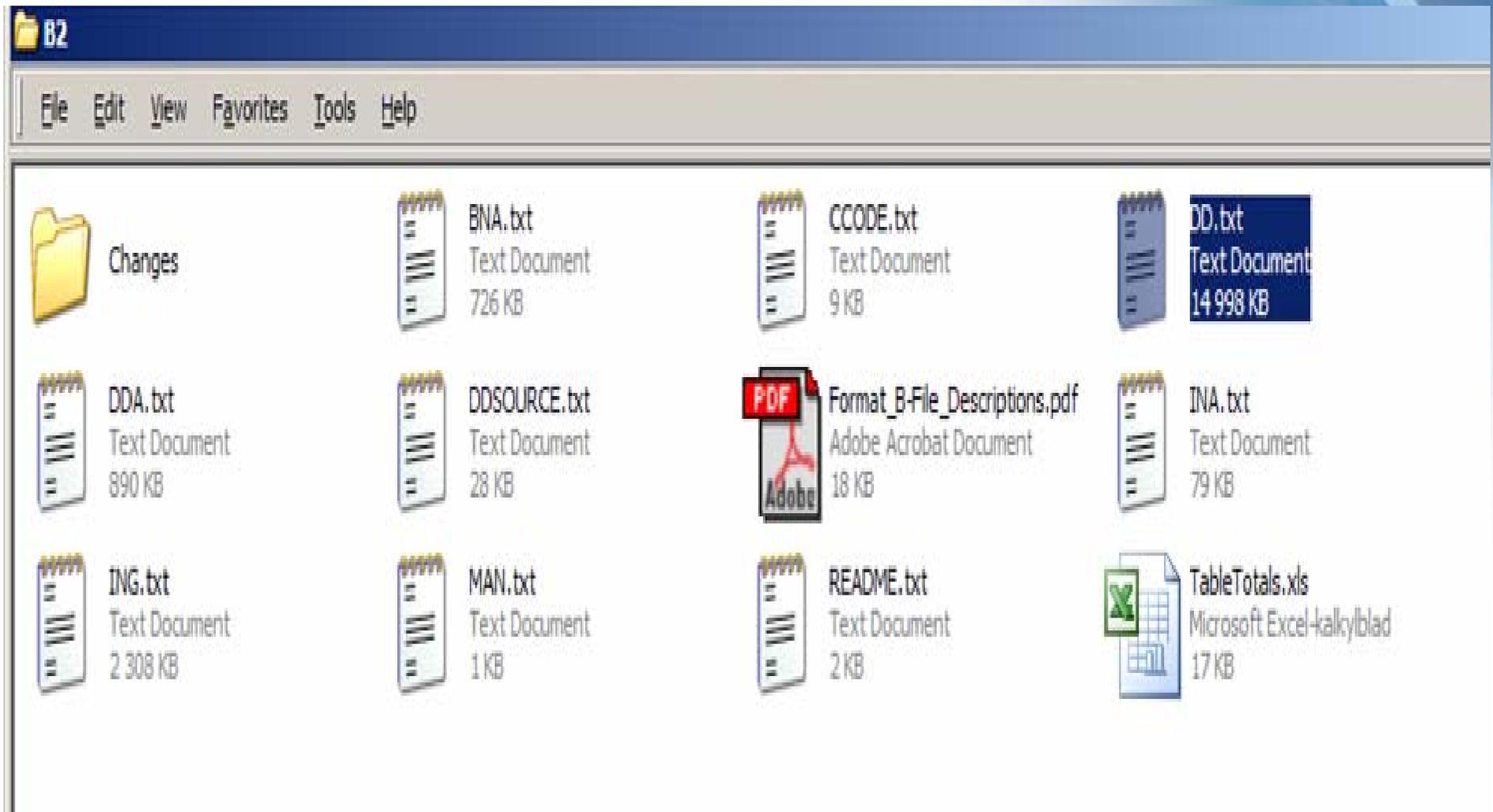
- The WHO Drug Dictionaries are available in different formats
- The formats are data-files, with pre-defined data-fields and relationships between the tables
- The data-files are loaded into the TMS



WHODrug Dictionary the B-2 Format



B-2 - files



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



DD table

```

DD.txt - Notepad
File Edit Format View Help
000001010016N INN 01 854METHYLDOPA
000001010024T04 124UNS 01 041ALDOMET /00000101/
000001010032T67 05 01 044PRESINOL
000001010040T68 09 01 044DOPAMET
000001010057T69 08 01 044HYPERPAXA
000001010065T69 06 01 044SEMBRINA
000001010073T70 25 01 044HYPERPAX
000001010081T73 38 01 044ALDOMIN
000001010099T77 46 01 044MEDOPA
000001010107T78 28 01 044MEDOPAL
000001010115T78 36 01 044POLINAL
000001010123T78 24 01 044METYLDOPA
000001010131T77 44 01 044DOPEGIT
000001010149T79 56 01 044NOVOMEDOPA
000001010156T81 51 01 044LEVOMET /00000101/
000001010164T81 02 01 044NOVOLEXIN
000001010172T76 36 01 044ELDOPANE
000001010180T82 69 01 044TENSIPAS
000001010198T83 02 01 044APO-METHYLDOPA
000001010206T82 71 01 044MEDOMET
000001010222T85 05 01 044ALDOMETIL
000001010230T86 21 01 044DOPAGEN
000001010248T86JPN 01 044MEDOBA
000001010255T73 36 01 044DESTENS
000001010263T86 64 01 044DIMAL
000001010271T86 21 01 044MELDOPA
000001010289T86 77 01 044DOPEGYT
000001010297T86 43 01 044EQUIBAR /00000101/
000001010305T89 19 01 044BAYPRESOL
000001010321T90 21 01 044ELANPRES
000001010339T92 101 01 044DANSUL
000001010347T91 49 01 044HYDOPA
000001010354T97 19 01 044ALDOPREN
000001010362T97 106 01 044HY-PO-TONE
000001010370T98JPN 01 044NICHIDOPAJO
000001010388T98 19 01 044METALPHA
000001010396T97 138 01 044EMDOPA
000001010412T01 132 01 044ALFAMETILDOPA
000001010420T01 132 01 044DOPANORE
000001010438T00 02 01 044NU-MEDOPA
000001010446T02 154 01 044ALMEDOP
000001010453T03 19UNS 01 041CARDIN /00000101/
000001010461N04 19 01 041METHYLDOPUM
000001010479T05 237UNS 01 051MEDOPASAL
000001010495T05 237UNS 01 051BECANTA
000001010503T05 237UNS 01 051METHYLDOPA TSURUHARA
000001010511T05 237UNS 01 051DANSUL NIPPON YAKKO
000001010529T05 237UNS 01 051DATTORYL
000001010537T05 237UNS 01 051DECABOLAMIN
000001010545T05 237UNS 01 051DES TENS
000001010552T05 237UNS 01 051DOMESIN
000001010560T05 237UNS 01 051DOPAMIN /00000101/
000001010578T05 237UNS 01 051GROSPISK
000001010586T05 237UNS 01 051HALFDOPA
000001010594T05 237UNS 01 051HYPERTEN /00000101/
000001010602T05 237UNS 01 051LEDOPAN
000001010610T05 237UNS 01 051MEPRIN
000001010628T05 237UNS 01 051METHOPLAIN
000001010636T05 237UNS 01 051U PRESS DOPA
000001010644T05 237UNS 01 051METHYLDOPA SCHERING
    
```



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



B-2 – file description

File descriptions WHO Drug Dictionary - Format B (B1 and B2)

Many fields in the dictionary are codes where the corresponding text can be found in the Subsidiary Files. These entries are identified by texts in *italics* in the Notes column. E.g *DDSOURCE.SOURCE_CODE*. Where *DDSOURCE* is the table and *SOURCE_CODE* is the field where the corresponding code can be found.

MAIN FILES

	Field	No of positions	Position in record	Notes	
DRUG DICTIONARY (DD)					
	Drug Record Number	1	6	1 - 6	Alphanumeric with leading zeroes
	Sequence Number 1	2	2	7 - 8	Alphanumeric with leading zeroes
	Sequence Number 2	3	3	9 - 11	Alphanumeric with leading zeroes
	Check Digit	4	1	12	Each Drug Code (Drecno, seq1, seq2) is assigned a check digit
	Designation	5	1	13	A one letter code classifying the type of a Drug Name
	Source Year	6	2	14 - 15	Denotes the year of issue (last two digits) of references source
	Source Code	7	4	16 - 19	Source of information. For non-proprietary names

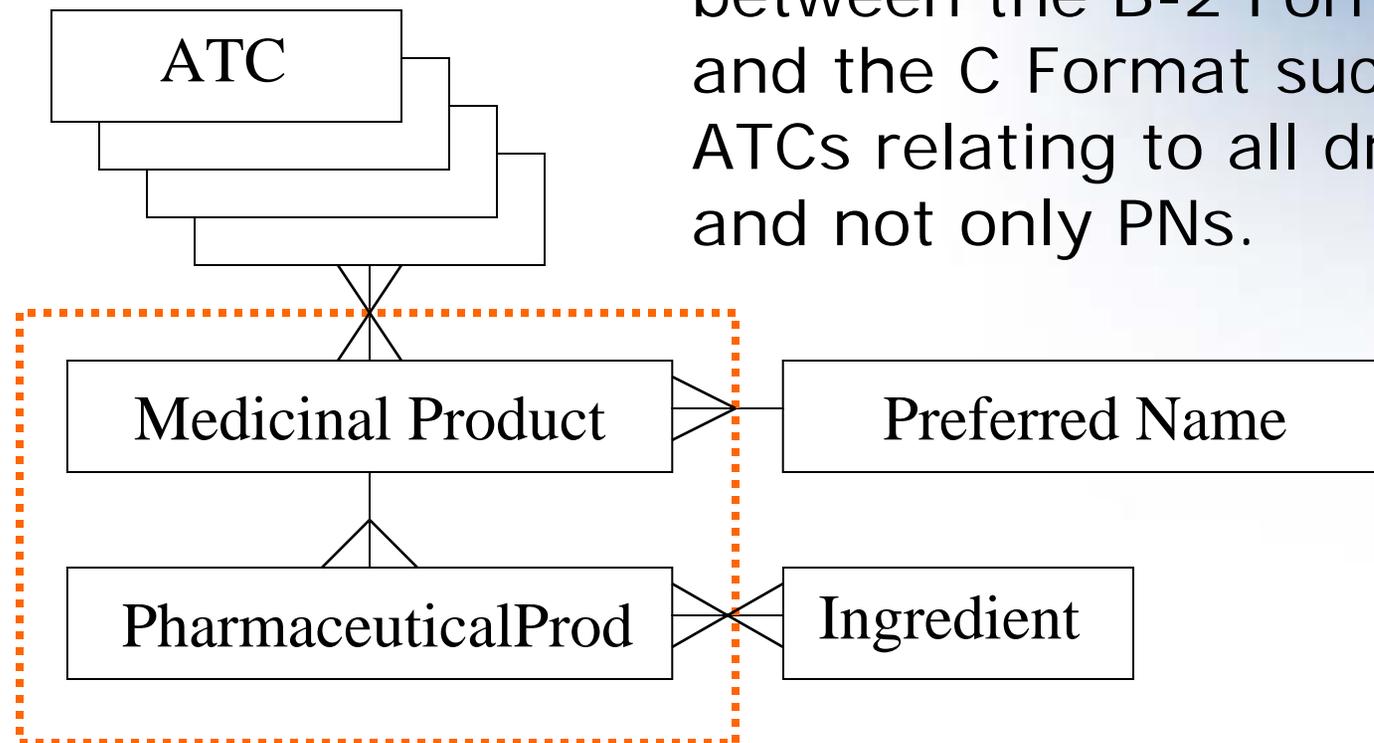


Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh



WHODrug Dictionary C Format

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

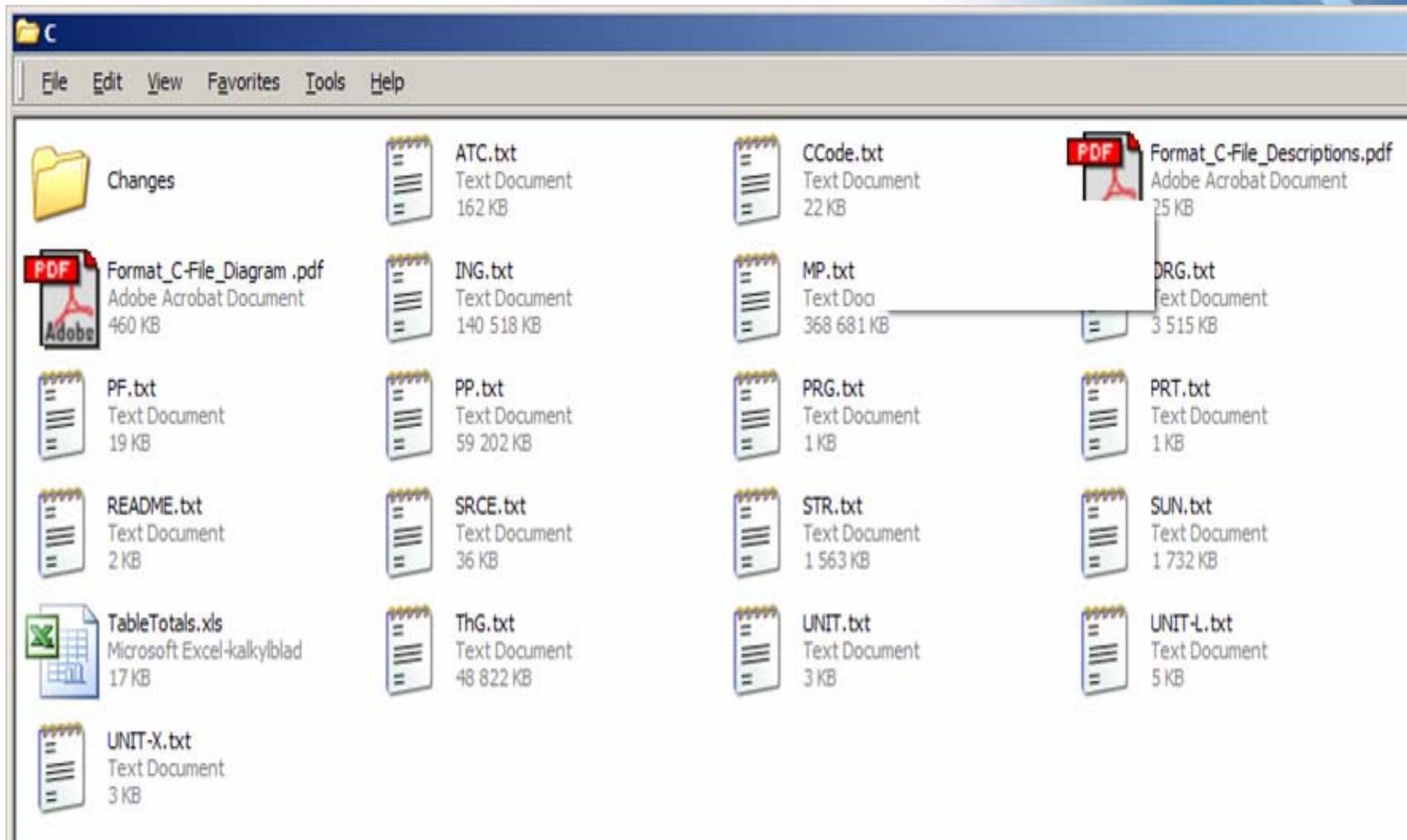


Overview of the C Format

- 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



C - files

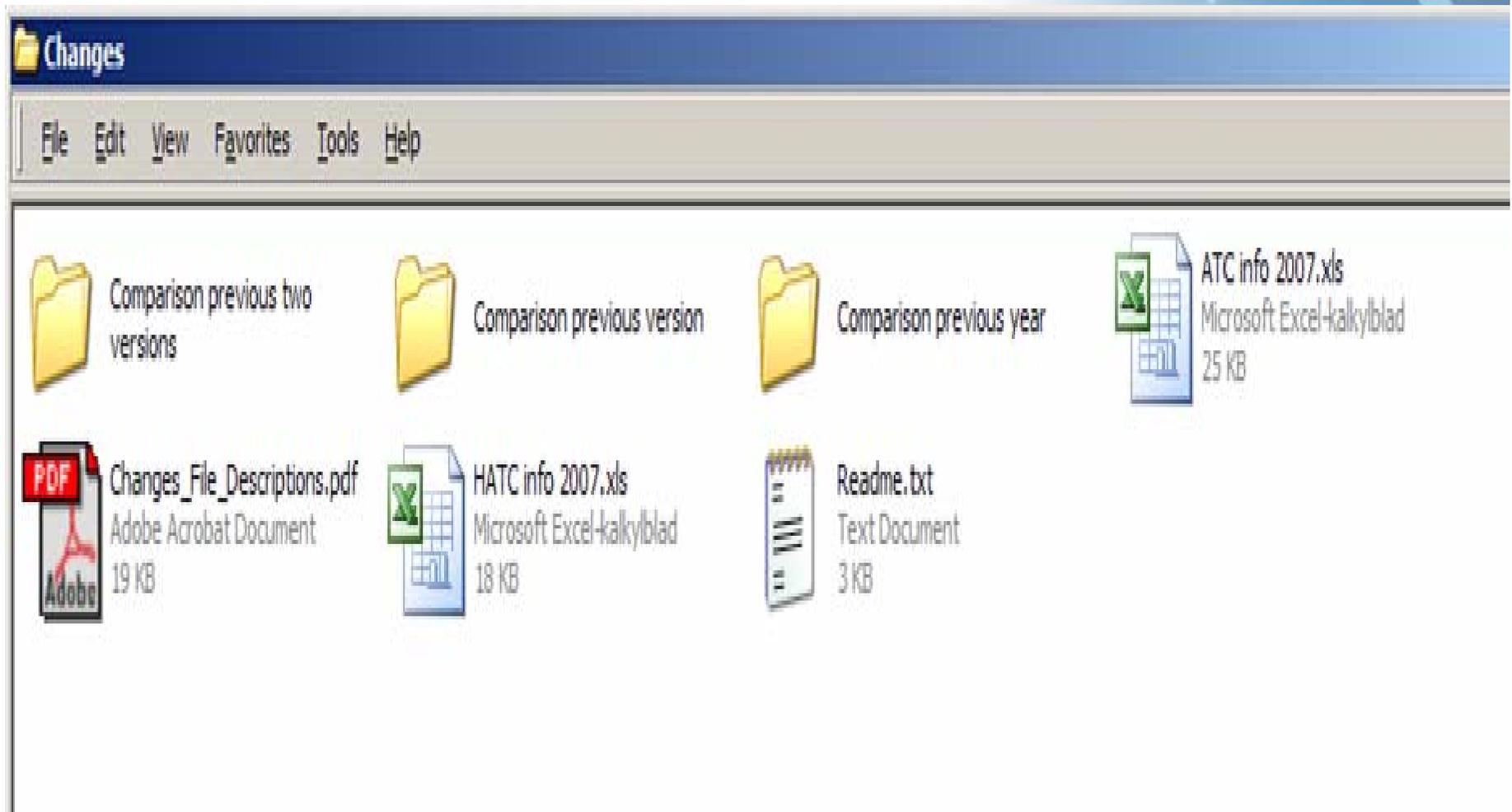


Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Changes files



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Why two formats?

- The B-2 format (previously known as the A format) has been in use over 20 years
- The C format was introduced in 2002
- The additional features in the C format is useful to:
 - Code with higher precision
 - Understand the difference between drugs with similar drug names



B and C Formats

- The B format is a dictionary of product names
 - Unique identifier – Drug Code (B-2)
- The C format is a dictionary of medicinal products. Each drug name can appear many times – e.g. in different forms and countries
 - Unique identifier – Medicinal Product ID
 - Drug Code is also included



B-1 and B-2

- The B format is available in two versions – B-1 and B-2.
 - In B-1 a name and a drug code can be repeated
 - In B-2 the name and the drug code is unique
- B-2 is the most commonly used dictionary format – B-1 is used by some companies as a reference



B-2 view

User: ddeemoo
Organisation: DEMO
Version: June 1, 2006
Dictionary: WHODDE
Fri, Sep 15, 2006
GMT: 14:48

[Search page](#)
[ATC tree](#)
Condensed
[Result](#)
[Product page](#)

Query: Product Name like alved

<u>Product Name</u>	<u>Drug Code</u>	<u>Ingredient(s)</u>
Alvedon	00020001004	Paracetamol
Alvedon dos	00020001337	Paracetamol

Number of rows: 2

Copyright © 2006 the Uppsala Monitoring Centre



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



C view

User: ddeemmo	Organisation: DEMO	Version: June 1, 2006	Dictionary: WHODDE	Fri, Sep 15, 2006 GMT: 14:49
------------------	-----------------------	--------------------------	-----------------------	---------------------------------

[Search page](#)
[ATC tree](#)
[Condensed](#)
[Result](#)
[Product page](#)

Query: Product Name like alved
Selection: Drug Code = 00020001004

<u>MP id</u>	<u>Drug Code</u>	<u>Product Name</u>	<u>Name specifier</u>	<u>Pharmaceutical Form</u>	<u>Strength</u>	<u>Country</u>	<u>MAH</u>
 188671	00020001004	Alvedon		SUPPOSITORIES, PAEDIATRIC	Unspecified	United Kingdom	Novex pharma
 188672	00020001004	Alvedon		Unspecified	Unspecified	United Kingdom	Novex pharma
 4765	00020001004	Alvedon		Unspecified	Unspecified	Sweden	Draco ab
 4967	00020001004	Alvedon	Forte	Unspecified	Unspecified	Sweden	Astra pharmaceutical products, inc.
 807830	00020001004	Alvedon		TABLETS	Unspecified	Philippines	Multicare pharm
 807831	00020001004	Alvedon		Unspecified	Unspecified	Philippines	Multicare pharm
 812839	00020001004	Alvedon		LIQUIDS	Unspecified	Philippines	Multicare pharm
 87552	00020001004	Alvedon		Unspecified	Unspecified	Unspecified	Not specified

Number of rows: 8

Copyright © 2006 the Uppsala Monitoring Centre



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



ICH compatible C format

- The International Conference of Harmonisation (ICH) is in the process of producing guidelines for medicinal product information.
- A number of fields were added in 2005 added, and some field lengths were expanded



Sequence 3 and 4

- Information about Pharmaceutical Form and Strength have been added to the Medicinal Product table
 - Sequence Number 3 – Pharmaceutical Form
 - Sequence Number 4 – Strength
- Facilitates the use of the C format, all relevant information is available in the same table



Use of Sequence 3 and 4

- With the additional fields all important datafields can be accessed in the Medicinal Product table – a ‘one table’ dictionary can be created.

MP_Id	Drug Name	Name Specifier	Drug Code	MaHolder	Country	Form	Strength	Ingredient
59142	Seresta	Forte Tabletten	000409 01 005	AHP AG	Switzerland	Tablets	50 Milligram	Oxazepam
59138	Seresta	Tabletten	000409 01 005	AHP AG	Switzerland	Tablets	15 Milligram	Oxazepam



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Information levels

	Drug Code	Name	Name specifier	Country	MAH	Form	Strength
1	X	X					
2	X	X	(X)				
3	X	X	(X)	X			
4	X	X	(X)	X	X		
5	X	X	(X)	X	X	X	
6	X	X	(X)	X	X	X	X



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



C format – information levels

MP_Id	Drug Name	Name Specifier	Drug Code	MaHolder	Country	Form	Strength	Ingredient
59142	Seresta	Forte Tabletten	000409 01 005	AHP AG	Switzerland	Tablets	50 Milligram	Oxazepam
59138	Seresta	Tabletten	000409 01 005	AHP AG	Switzerland	Tablets	15 Milligram	Oxazepam
59139	Seresta	Tabletten	000409 01 005	AHP AG	Switzerland	Tablets		Oxazepam
59140	Seresta		000409 01 005	AHP AG	Switzerland			Oxazepam
405769	Seresta		000409 01 005	Biodim	France	Tablets		Oxazepam
405770	Seresta		000409 01 005	Biodim	France			Oxazepam
8477	Seresta		000409 01 005	Wyeth	Netherlands			Oxazepam
59141	Seresta		000409 01 005					Oxazepam



Non-unique names

- Some drug names can mean many things – the names can be used in different countries or forms with different active ingredients
- In the B-2 format the Drug Record number and Sequence number 1 is added to the drug name – to make it unique
- In the C format entries have additional data fields



Non-unique name, B-2 format

ACTRON	/00020001/
ACTRON	/00109201/
ACTRON	/00321701/
ACTRON	/00391201/
ACTRON	/00727101/



Non-unique name, C format

Drug Name	Name Specifier	Drug Code	MaHolder	Country	Ingredients
Actron		003912 01 026	Bayer	France	Acetylsalicylic acid Caffeine Paracetamol
Actron		001092 01 461	Bayer	Mexico	Ibuprofen
Actron		003217 01 053	Bayer	United States	Ketoprofen
Actron		000200 01 158	Bayer	Spain	Paracetamol
Actron	/Old form/	007271 01 001	Miles Martin	Spain	Acetylsalicylic acid Caffeine Citric acid Paracetamol Sodium bicarbonate



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.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



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).



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



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



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh



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.



Part III: Loading and Updating the B2 and C Formats into TMS



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Deriving ATC Codes from WHODrug in TMS to OC

- Since TMS requires a Single Derivable Path to derive dictionary terms to an External System (such as AERS or OC), Drug Names with Multiple ATCs can NOT send ALL possible ATCs to OC.
- This problem occurs regardless of WHODrug Dictionary Format. In B2 Format, it occurs for Preferred Name (Generic) Drugs only, but in C Format, it occurs for ALL Drugs.
- 4 Common options for dealing with this situation in TMS follow



Loading ATC codes in Type B2 and Type C

- Option 1: Concatenating ATC codes as level detail of Preferred Name or Drug Name.
 - Since the Drug Name is always derivable, the entire set of ATC codes becomes a concatenated string, which is a Level Detail Attribute of the Drug Name or Preferred Term.
 - This requires parsing of the concatenated ATC Codes within OC by Derivation Procedures, or within SAS.



Loading ATC codes in Type B2 and Type C (2)

- Option 2: Create a Primary link to the ATC codes based on some programmatic rule defined by the business users or with a "MULTIPLE" flag
 - Could be based on common occurrences of ATCs, known indications, or even alphabetical order although this is not recommended
- In addition also set a "MULTIPLE" ATC or Level Detail which would indicate to an OC Data Manager that multiple ATCs were possible and therefore, High-Level Reclassification might be necessary. Without this MULTIPLE indicator, a strong knowledge of ATC classifications would be required at the OC level to know whether or not multiple ATCs were possible.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Loading ATC codes in Type B2 and Type C (2)

- Option 3: Load Separate Drug and ATC Dictionaries.
 - Loading Drug Names into a first dictionary and ATCs into a second dictionary
 - The 2nd ATC Dictionary would have a classification level as the concatenation of the Preferred Drug Name and ATC code
 - A derivation procedure populates the VT level of this 2nd ATC dictionary from the classified Preferred Drug name (from the first dictionary) concatenated with Indication or Route for coding in the 2nd dictionary.
 - This requiring two Batch Validations, which is sometimes called a "split" WHODrug dictionary solution.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited

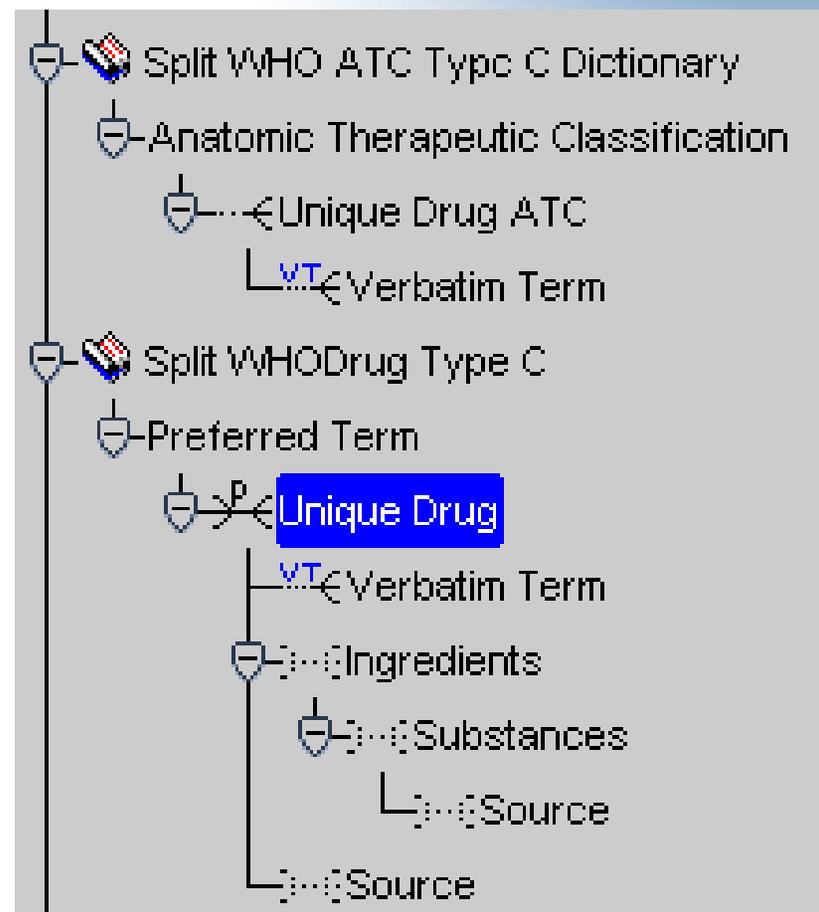
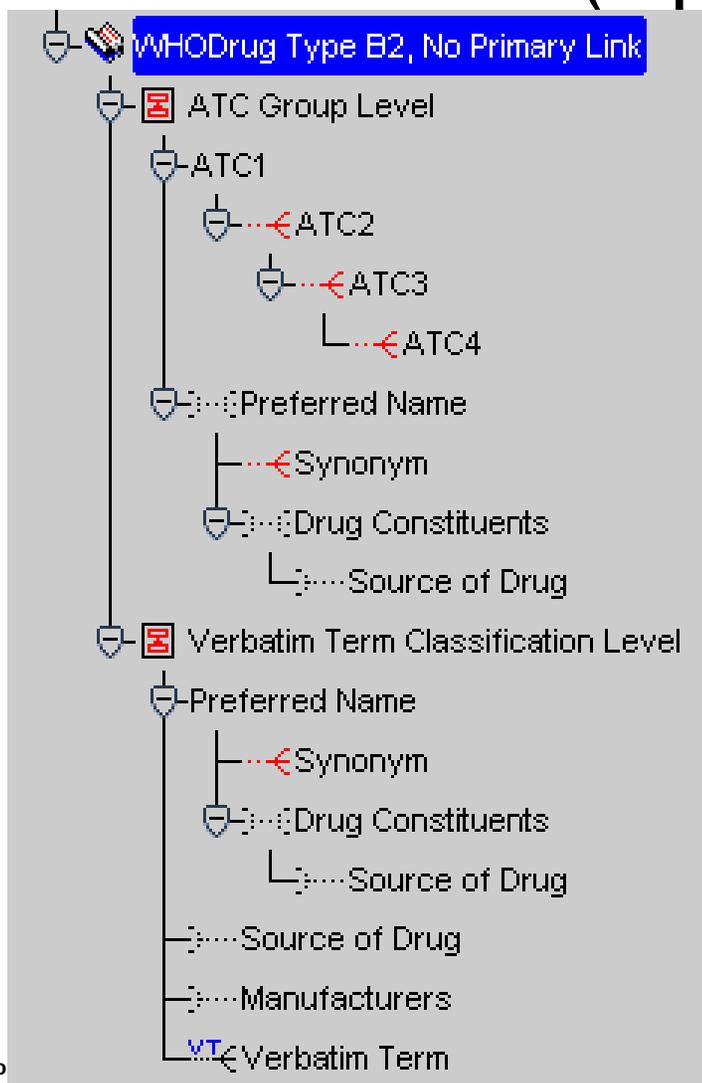


Loading ATC codes in Type B2 and Type C (3)

- Option 4: Do not derive ATC codes and create a custom view for retrieving all ATC codes into SAS.



View of TMS, WHODrug B2 without a PL (Option 1) and WHODrug Split Dictionary (Option 3)



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.



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 the 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.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



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.



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

The screenshot displays a hierarchical tree structure on the left side of the interface. The root node is 'Preferred Term', which branches into 'Unique Drug' (highlighted in blue), 'Verbatim Term', 'Ingredients', 'Substances', and 'Source'. The 'Unique Drug' node further branches into 'Ingredients', 'Substances', and 'Source'. The 'Substances' node branches into 'Source'. The right side of the interface shows a query results table with a 'Query' dropdown set to 'Standard'. The table contains the following data:

Unique Drug	
T	AMPICILLIN 275
T	AMPICILLIN 247

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

Unique Drug

Unique Drug	Level	Medicinalproc	Sequence_key
T BENADRYL WARNER LAMBERT CONSUMER HEALTH IRL	UNIQUEDF	52370	00647601002
T BENADRYL WARNER LAMBERT CONSUMER HEALTH USA	UNIQUEDF	51457	00000402049
T BENADRYL WARNER LAMBERT DNK	UNIQUEDF	52616	00945501004
T BENADRYL WARNER LAMBERT ESP	UNIQUEDF	51459	00000402051
T BENADRYL WARNER LAMBERT GBR	UNIQUEDF	52615	00945501003
T BENADRYL WARNER LAMBERT HRG	UNIQUEDF	51462	00000402054
T BENADRYL WARNER LAMBERT ITA	UNIQUEDF	52393	00673901009
T BENADRYL WARNER LAMBERT THA	UNIQUEDF	51458	00000402050

Relation	Term	RGlB?	Appr?	Alt Code	Type
T Strong	SODIUM CITRATE 38 84982	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary
_ T	MENTHOL 38 84983	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary
_ T	DIPHENHYDRAMINE HYDROCHLORIDE 38 84984	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary
_ T	AMMONIUM CHLORIDE 38 84985	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary
T	MARTINDALE - THE COMPLETE DRUG REFERENC	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Same Drug Name in Different Countries (2)

Benadryl in the United Kingdom

Query Standard

Unique Drug	Level	Medicinalproc	Sequence_k
T BENADRYL WARNER LAMBERT CONSUMER HEALTH IRL	UNIQUEDF	52370	00647601002
T BENADRYL WARNER LAMBERT CONSUMER HEALTH USA	UNIQUEDF	51457	00000402049
T BENADRYL WARNER LAMBERT DNK	UNIQUEDF	52616	00945501004
T BENADRYL WARNER LAMBERT ESP	UNIQUEDF	51459	00000402051
T BENADRYL WARNER LAMBERT GBR	UNIQUEDF	52615	00945501003
T BENADRYL WARNER LAMBERT HKO	UNIQUEDF	51462	00000402054
T BENADRYL WARNER LAMBERT ITA	UNIQUEDF	52393	00673901009
T BENADRYL WARNER LAMBERT THA	UNIQUEDF	51458	00000402050

Relation	Term	RGlB?	Appr?	Alt Code	Type
T Strong	ACRIVASTINE 38 85279	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary
_ T Strong	MARTINDALE - THE COMPLETE DRUG REFERENC	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		Dictionary



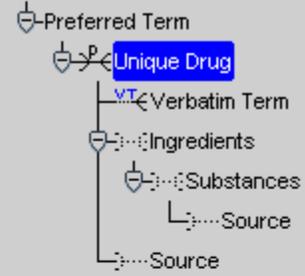
Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Changes in the Drug

Robitussin AC



Entered in 2002

Unique Drug					
	Level	Medicinalproc	Sequence_ke	SubType	
T	ROBITUSSIN A-C /OLD FORM/ ROBINS A.H. COMPANY, INCORPORA	UNIQUEDF	11947	00074201001	Company
T	ROBITUSSIN AC ROBINS A.H. COMPANY, INCORPORATED USA COD	UNIQUEDF	35354	00693301008	Company
T	ROBITUSSIN AC WHITEHALL-ROBINS INC. CAN CODEINE PHOSPHA1	UNIQUEDF	11948	00074201002	Company

Relation	Term	Level	Code	RGlb?	Appr? A
T Strong	CODEINE PHOSPHATE 38 51900	INGWHO03Q4-ING	51900	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
T Strong	GUAIFENESIN 38 51901	INGWHO03Q4-ING	51901	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
T Strong	AMERICAN DRUG INDEX	INGWHO03Q4-SRCE	010	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Changes in the Drug (2)

Robitussin AC

⊖ Dictionaries

- ⊖ Ingenix_WHOATC_03Q4
- ⊖ Ingenix_WHODrug_03Q4
 - ⊖ Preferred Term
 - ⊖ Unique Drug
 - VT Verbatim Term
 - ⊖ Ingredients
 - ⊖ Substances
 - ⊖ Source
 - ⊖ Source

Query Standard

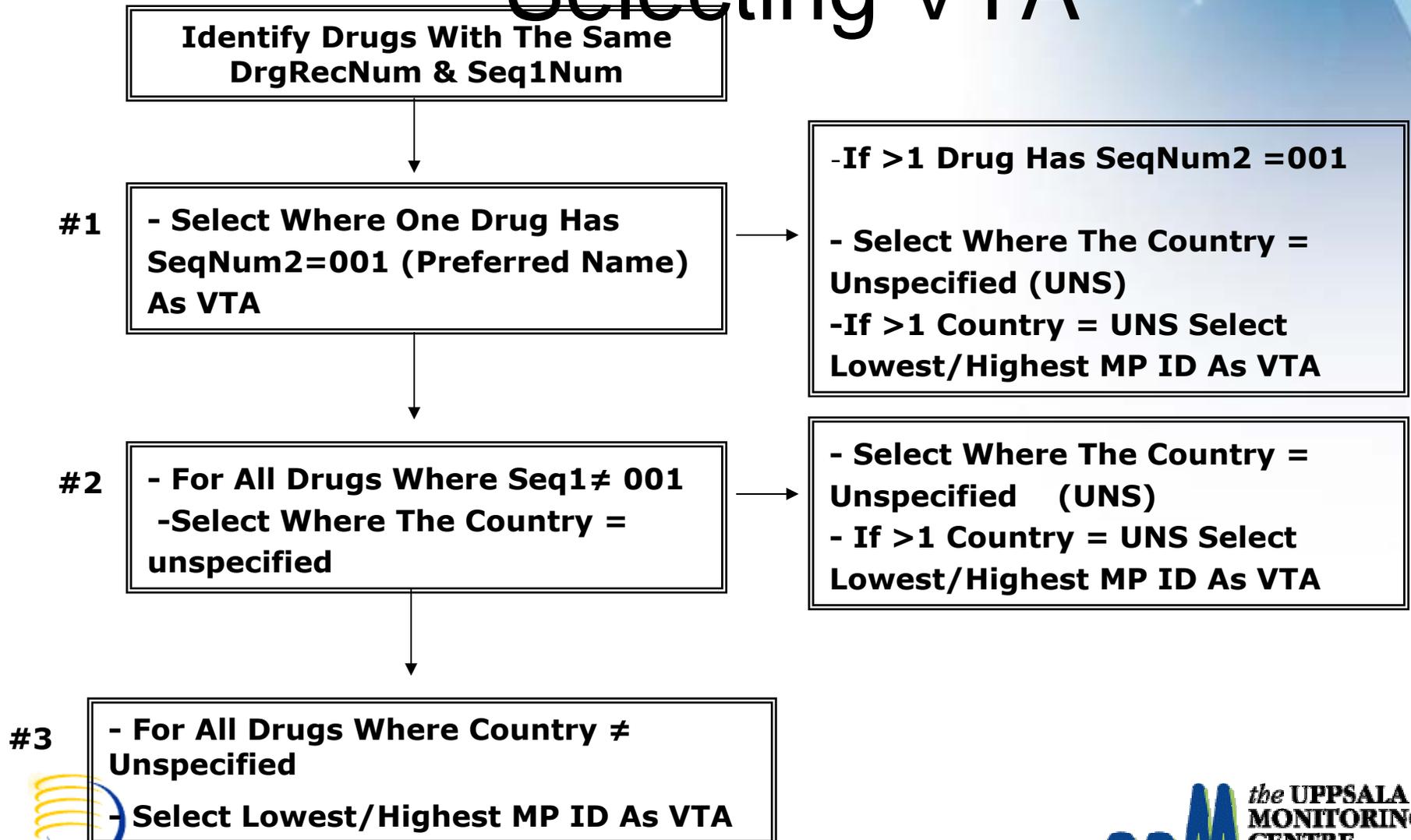
Unique Drug	Level	Medicinalproc	Sequence_key	Su
ROBITUSSIN A-C OLD FORM/ROBINS A.H. COMPANY, INCORPORA	UNIQUEDF	11947	00074201001	C
ROBITUSSIN AC ROBINS A.H. COMPANY, INCORPORATED USA COE	UNIQUEDF	35354	00693301008	C
ROBITUSSIN AC WHITEHALL-ROBINS INC. CAN CODEINE PHOSPHAT	UNIQUEDF	11948	00074201002	C

Relation	Term	Level	Code	RGlb?
T Strong	CODEINE PHOSPHATE 38 13807	INGWHO03Q4-ING	13807	✓
_ T Strong	GUAIFENESIN 38 13808	INGWHO03Q4-ING	13808	✓
_ T Strong	PHENIRAMINE MALEATE 38 13809	INGWHO03Q4-ING	13809	✓
_ T Strong	AMERICAN DRUG INDEX	INGWHO03Q4-SRCE	010	✓
_ T Strong	ROBITUSSIN A-C	INGWHO03Q4-VT		✓

Entered in 1985



A Suggested Decision Tree for Selecting VTA



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.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



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.



Loading and Update Considerations for B2 Format

- Since the Drug Recnum + Sequence 1 have been added to the B2 format for Drug Names which have multiple Drug Record Numbers, some Drug Names which previously autocoded do not currently autocode.
- While this represents a small percentage of Drug Names in quantitative terms, these drugs are the most commonly used and therefore occur the most frequently



Why does Aspirin No Longer Autocode?

- Consider the drug aspirin in the WHODD or WHODDE B2 format dictionary:
 - In the case of WHODD, the single occurrence of aspirin appears with a drug record number appended, to indicate that other drug record numbers are possible
 - In the case of WHODDE, multiple occurrences of aspirin exist with different drug record numbers



Possible Solutions

- Use a similar algorithm for WHODrug Type C format loading for B2.
 - Requires establishing domain VTA rules for each of the multiple sets of Drug Recnums
 - Drug Names could be defaulted based on country or Preferred Name derivation
 - Create Global VTAs where a single drug exists with a Drug Code appended if the WHODD Type is being used.
- If the goal of coding is ONLY to derive Preferred Names and NOT ATCs, then it is possible to create a Global VTA if all the Preferred Names are the same, even if the Drug Recnums are different
- Possible enhancements to TMS to allow “single” VTA coding (formerly called VTI functionality), which is similar to HLC at the VT coding level instead
- Derive a specific match based on Site/Investigator/Patient location or country, and use this in a derived question or Search Object.



Authors: Anders Hansson, Daniel von Sydow, Sunil G. Singh

©2007 DBMS Consulting, Inc. and Uppsala Monitoring Centre: Unauthorized Duplication is Strictly Prohibited



Questions?

- Write to the UMC:
 - Daniel.vonsydow@umc-products.org
 - Anders.hansson@umc-products.org
- Write to DBMS Consulting:
 - singh@clinicalserver.com

