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
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.
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
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
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
### DD table

<table>
<thead>
<tr>
<th>ID</th>
<th>INN</th>
<th>UNR</th>
<th>Format</th>
<th>View</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000100103161</td>
<td>01</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001002474</td>
<td>124</td>
<td>UNS</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001002736</td>
<td>05</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001004018</td>
<td>08</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001005779</td>
<td>08</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001005779</td>
<td>06</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001007779</td>
<td>25</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001008172</td>
<td>28</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001009977</td>
<td>46</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001010778</td>
<td>28</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001012578</td>
<td>36</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001012378</td>
<td>24</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001013177</td>
<td>46</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001014379</td>
<td>56</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001015678</td>
<td>51</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001016478</td>
<td>02</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001017276</td>
<td>36</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001018578</td>
<td>69</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001019873</td>
<td>02</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001020678</td>
<td>71</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001022778</td>
<td>05</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001023078</td>
<td>21</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001025816</td>
<td>JPN</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001025873</td>
<td>36</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001026878</td>
<td>04</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001027178</td>
<td>21</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001028778</td>
<td>77</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001029786</td>
<td>13</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001030779</td>
<td>19</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001032170</td>
<td>21</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001033292</td>
<td>101</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001034718</td>
<td>49</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001035479</td>
<td>19</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001036279</td>
<td>106</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001037978</td>
<td>JPN</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001038878</td>
<td>19</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001039687</td>
<td>138</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001043170</td>
<td>132</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001042501</td>
<td>212</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001043870</td>
<td>02</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001044600</td>
<td>134</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001045370</td>
<td>193</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001046104</td>
<td>193</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001047903</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001049503</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001050370</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001051170</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001052070</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001053760</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001054570</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001055205</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001056070</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001057803</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001059870</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001062870</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001061070</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001063670</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>000001001064470</td>
<td>287</td>
<td>4541MDL</td>
<td>/000010101/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 italic 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**

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

Query: Product Name like alved

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Drug Code</th>
<th>Ingredient(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvedon</td>
<td>00020001004</td>
<td>Paracetamol</td>
</tr>
<tr>
<td>Alvedon dos</td>
<td>00020001337</td>
<td>Paracetamol</td>
</tr>
</tbody>
</table>

Number of rows: 2
C view

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

Number of rows: 8
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.

<table>
<thead>
<tr>
<th>MP_Id</th>
<th>Drug Name</th>
<th>Name Specifier</th>
<th>Drug Code</th>
<th>MaHolder</th>
<th>Country</th>
<th>Form</th>
<th>Strength</th>
<th>Ingredient</th>
</tr>
</thead>
<tbody>
<tr>
<td>59142</td>
<td>Seresta</td>
<td>Forte Tabletten</td>
<td>000409 01 005</td>
<td>AHP AG</td>
<td>Switzerland</td>
<td>Tablets</td>
<td>50 Milligram</td>
<td>Oxazepam</td>
</tr>
<tr>
<td>59138</td>
<td>Seresta</td>
<td>Tabletten</td>
<td>000409 01 005</td>
<td>AHP AG</td>
<td>Switzerland</td>
<td>Tablets</td>
<td>15 Milligram</td>
<td>Oxazepam</td>
</tr>
</tbody>
</table>
## Information levels

<table>
<thead>
<tr>
<th></th>
<th>Drug Code</th>
<th>Name</th>
<th>Name specifier</th>
<th>Country</th>
<th>MAH</th>
<th>Form</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### C format – information levels

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

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Name Specifier</th>
<th>Drug Code</th>
<th>MaHolder</th>
<th>Country</th>
<th>Ingredients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actron</td>
<td></td>
<td>003912 01 026</td>
<td>Bayer</td>
<td>France</td>
<td>Acetylsalicylic acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caffeine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Paracetamol</td>
</tr>
<tr>
<td>Actron</td>
<td></td>
<td>001092 01 461</td>
<td>Bayer</td>
<td>Mexico</td>
<td>Ibuprofen</td>
</tr>
<tr>
<td>Actron</td>
<td></td>
<td>003217 01 053</td>
<td>Bayer</td>
<td>United States</td>
<td>Ketoprofen</td>
</tr>
<tr>
<td>Actron</td>
<td></td>
<td>000200 01 158</td>
<td>Bayer</td>
<td>Spain</td>
<td>Paracetamol</td>
</tr>
<tr>
<td>Actron</td>
<td>/Old form/</td>
<td>007271 01 001</td>
<td>Miles Martin</td>
<td>Spain</td>
<td>Acetylsalicylic acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caffeine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Citric acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Paracetamol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sodium bicarbonate</td>
</tr>
</tbody>
</table>
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: 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.
Part III: Loading and Updating the B2 and C Formats into TMS
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.
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.
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 cannot 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.
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 XXXXXXX01001. 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

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: AutoEncoding

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

<table>
<thead>
<tr>
<th>Unique Drug</th>
<th>Level</th>
<th>Medicinalproc</th>
<th>Sequence_key</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENADRYL WARNER LAMBERT CONSUMER HEALTH IRL</td>
<td>UNIQUEDF</td>
<td>52370</td>
<td>00647601002</td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT CONSUMER HEALTH USA</td>
<td>UNIQUEDF</td>
<td>51457</td>
<td>00000402049</td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT DNK</td>
<td>UNIQUEDF</td>
<td>52616</td>
<td>00945601004</td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT ESP</td>
<td>UNIQUEDF</td>
<td>51459</td>
<td>00000402051</td>
</tr>
<tr>
<td><strong>BENADRYL WARNER LAMBERT GBR</strong></td>
<td><strong>UNIQUEDF</strong></td>
<td><strong>52615</strong></td>
<td><strong>00945601003</strong></td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT HKG</td>
<td>UNIQUEDF</td>
<td>51462</td>
<td>00000402054</td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT ITA</td>
<td>UNIQUEDF</td>
<td>52393</td>
<td>00673301009</td>
</tr>
<tr>
<td>BENADRYL WARNER LAMBERT THA</td>
<td>UNIQUEDF</td>
<td>51458</td>
<td>00000402050</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation</th>
<th>Term</th>
<th>R格?</th>
<th>Appl?</th>
<th>Alt Code</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Strong</td>
<td>ACRIVASTINE 38 65279</td>
<td>✔️</td>
<td>✔️</td>
<td>Dictionary</td>
</tr>
<tr>
<td>_T</td>
<td>Strong</td>
<td>MARTINDALE - THE COMPLETE DRUG REFERENCE</td>
<td>✔️</td>
<td>✔️</td>
<td>Dictionary</td>
</tr>
</tbody>
</table>
Changes in the Drug
Robitussin AC

Entered in 2002
Changes in the Drug (2)

Robitussin AC

Entered in 1985
A Suggested Decision Tree for Selecting VTA

1. Identify Drugs With The Same DrgRecNum & Seq1Num
   - Select Where One Drug Has SeqNum2=001 (Preferred Name) As VTA
   - If >1 Drug Has SeqNum2 =001
     - Select Where The Country = Unspecified (UNS)
     - If >1 Country = UNS Select Lowest/Highest MP ID As VTA

2. For All Drugs Where Seq1≠ 001
   - Select Where The Country = unspecified
   - Select Where The Country = Unspecified (UNS)
   - If >1 Country = UNS Select Lowest/Highest MP ID As VTA

3. For All Drugs Where Country ≠ Unspecified
   - Select Lowest/Highest MP ID As 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 DICTCONTENT_CODE in the dictionary which comes from the vendor which can be compared with queries against the vendor source data to determine what DICTCONTENT_CODEs to insert/update/delete.
  - Additionally, during the dictionary load process, it is not required to specify a DICTCONTENT_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.
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.
Questions?

• Write to the UMC:
  – Daniel.vonsydow@umc-products.org
  – Anders.hansson@umc-products.org

• Write to DBMS Consulting:
  – singh@clinicalserver.com