FDA terminology for substances

Yulia Borodina, Larry Callahan, Frank Switzer
FDA/OC/OO/OITI/Informatics Standards and Scientific Computing
Outline

1. Basic facts about FDA’s Substance terminology
2. Substance Indexing initiative
3. Elements of Substance definition
Substance terminology facts

- The substance terminology has been created in response to the demand to uniquely identify ingredients (substances) in medicinal products for regulatory purposes.

- The substance terminology is being created by FDA Substance Registration System (SRS) at [http://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniQIngridIdentiferUNII/](http://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniQIngridIdentiferUNII/).

- Each substance registered in SRS has a **definition** – set of characteristics that unambiguously identifies the substance.

- Each substance is assigned a Unique Ingredient Identifier (UNII).
• Characteristics essential for **substance definition** are described in the ISO standard ISO /FDIS 11238 ISO/FDIS: Health Informatics - Identification of Medicinal Products Data (IDMP)

• IDMP standard distinguishes following types of substances:
  • chemical
  • protein
  • nucleic acid
  • polymer
  • structurally-diverse substance

• IDMP standard allows a substance to be a mixture of substances
  • “Mixtures consist of a simple combination of single substances that are either isolated together or are the result of the same synthetic process.”
• Companies submitting product information using Structured Product Labeling (SPL) use **UNII** to code the active and inactive ingredients of the product

• **UNII**s should be provided for substances/ingredients in drugs, biologics, medical foods and medical devices

• **UNII**s can be obtained through a name-to-UNII look-up at [http://fdasis.nlm.nih.gov/srs/](http://fdasis.nlm.nih.gov/srs/) or can be received through a request made to the FDA SRS group sent at [FDA-SRS@fda.hhs.gov](mailto:FDA-SRS@fda.hhs.gov)

• Over **60 000 UNII** containing SPL files submitted by companies are available for download at [http://dailymed.nlm.nih.gov/dailymed/downloadLabels.cfm](http://dailymed.nlm.nih.gov/dailymed/downloadLabels.cfm)
Example: ingredients listed in the drug label

<table>
<thead>
<tr>
<th>Active ingredient</th>
<th>UNII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dextromethorphan</td>
<td>7355X3ROTS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inactive ingredients</th>
<th>UNII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polistirex</td>
<td>5H9W9GTW27</td>
</tr>
<tr>
<td>citric acid monohydrate</td>
<td>2968PHW8QP</td>
</tr>
<tr>
<td>edetate disodium</td>
<td>7FLD91C86K</td>
</tr>
<tr>
<td>ethylcelluloses</td>
<td>7Z8S9VYZ4B</td>
</tr>
<tr>
<td>FD&amp;C Yellow No. 6</td>
<td>H77VEI93A8</td>
</tr>
<tr>
<td>high fructose corn syrup</td>
<td>XY6UN3QB6S</td>
</tr>
<tr>
<td>methylparaben</td>
<td>A2I8C7HI9T</td>
</tr>
<tr>
<td>polyethylene glycol 3350</td>
<td>G2M7P15E5P</td>
</tr>
<tr>
<td>polysorbate 80</td>
<td>6OZP39ZG8H</td>
</tr>
<tr>
<td>propylene glycol</td>
<td>6DC9Q167V3</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Substance Indexing

• Substance Indexing is a part of the SPL Indexing initiative which has the goal of enhancing access to the electronic product information provided by the companies. Indexing refers to the creation by FDA of one or more files with machine-readable annotations that can be linked to the product SPL provided by the company.

• Substance Indexing provides the FDA Substance terminology in an electronic format

• Substance Indexing enhances users' ability to automatically search and sort information on substances included in a product

• Substance Indexing enhances users' ability to select a UNII based on substance definition and not on names
Substance Index Files are in Structure Product Labeling format (SPL XML)* and are regularly maintained and updated by the FDA. Each Substance Index File includes versioning information to allow tracking of changes.

Substance Index Files can be downloaded from http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm377913.htm

Example of Product SPL File, Pharm Class Index File and Substance Index File

```xml
<code code="48780-1" displayName="SPL LISTING DATA ELEMENTS SECTION" codeSystem="2.16.840.1.113883.6.1"/>

<manufacturedProduct>
  <code code="54868-3690" codeSystem="2.16.840.1.113883.6.69"/>
  <name>Lotensin</name>
  <ingredient classCode="ACTIB">
    ...
  </ingredient>
  <ingredientSubstance>
    <code code="N1SN99T69T" codeSystem="2.16.840.1.113883.4.9"/>
    <name>BENAZEPRIL HYDROCHLORIDE</name>
    <activeMoiety>
      <code code="JRM708L703" codeSystem="2.16.840.1.113883.4.9"/>
      <name>BENAZEPRILAT</name>
    </activeMoiety>
  </ingredientSubstance>
</manufacturedProduct>

<code code="60685-5" codeSystem="2.16.840.1.113883.6.1" displayName="Indexing - Pharmacologic Class"/>

<identifiedSubstance>
  <code code="JRM708L703" codeSystem="2.16.840.1.113883.4.9"/>
  <name use="L">BENAZEPRILAT</name>
  <asSpecializedKind>
    <generalizedMaterialKind>
      <code code="N0000175562" codeSystem="2.16.840.1.113883.3.26.1.5" displayName="Angiotensin Converting Enzyme Inhibitor [EPC]"/>
      <name use="L">angiotensin converting enzyme inhibitor</name>
    </generalizedMaterialKind>
  </asSpecializedKind>
</identifiedSubstance>

<code code="64124-1" codeSystem="2.16.840.1.113883.6.1" displayName="Indexing - Substance"/>

<identifiedSubstance>
  <code code="JRM708L703" displayName="benazeprilat" codeSystem="2.16.840.1.113883.4.9"/>
  <characteristic>
    <code displayName="Chemical Structure" codeSystem="2.16.840.1.113883.3.26.1.1" code="C103240"/>
    <value p4:type="ED" mediaType="application/x-inchi">InChI=1S/C22H24N2O5/c25-20(26)14-24-19-9-5-4-8-16(19)11-13-17(21(24)27)23-18(22(28)29)12-10-15-6-2-1-3-7-15/h1-9,17-18,23H,10-14H2,(H,25,26)(H,28,29)/t17-,18-/m0/s1</value>
  </characteristic>
</identifiedSubstance>

< WARNING: FETAL TOXICITY: This drug may cause fetal harm when administered to pregnant women. Use only when clearly needed when the benefits outweigh the risks. >
```

CH₂COOH
```
Content of Substance Index File

- UNII
- Substance equivalence codes
- Substance names
- Substance definition
UNII

**identifiedSubstance**
- **code**
  - **displayName**
  - **codeSystem**
- **name**
- **asNamedEntity**
  - **code**
    - **displayName**
    - **codeSystem**
  - **name**
- **asEquivalentSubstance**
  - **definingSubstance**
    - **code**
      - **code**
      - **codeSystem**

**Definition hash code**

<table>
<thead>
<tr>
<th>UNII</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHS9PGQ2LS</td>
<td>METHYLEPHEDRINE, (+/-)-</td>
</tr>
<tr>
<td>2.16.840.1.113883.4.9</td>
<td></td>
</tr>
<tr>
<td>C43707</td>
<td>primary name</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.26.1.1</td>
<td></td>
</tr>
<tr>
<td>6a3ea792-4331-a4ba-7641-998a1e32c2bd</td>
<td></td>
</tr>
<tr>
<td>2.16.840.1.113883.3.2705</td>
<td></td>
</tr>
</tbody>
</table>
Elements of Substance Definition

- Substance has one or more structural units (moieties)
  - Moiety has a type
    - mixture component
    - chemical moiety
    - protein subunit
    - none (for single-moiety substances)
  - Moiety may have a structure
    - chemical structure
      - must have InChI
    - amino acid sequence
    - nucleobase sequence
  - Moiety may have sub-moieties
  - Moiety has a quantity
    - quantity may be undefined
    - quantity may be a range
**Example: single chemical compound**

<table>
<thead>
<tr>
<th>moiety</th>
<th>numerator</th>
<th>denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>quantity</td>
<td>value</td>
<td>value</td>
</tr>
<tr>
<td></td>
<td>unit</td>
<td>unit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>partMoiety</th>
<th>code</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>code</td>
<td>unit</td>
</tr>
<tr>
<td></td>
<td>codeSystem</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>subjectOf</th>
<th>characteristic</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p4:type</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mediaType</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xmlns:p4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>#text</td>
</tr>
</tbody>
</table>

```
1
1
1
1

OMFE10J96E
2.16.840.1.113883.4.9

Chemical Structure
2.16.840.1.113883.3.26.1.1
C103240

ED
application/x-inchi
http://www.w3.org/2001/XMLSchema-instance
InChI=1S/H2O4S.Sn/c1-5(2,3)4;/h(H2,1,2,3,4);/q:+2/p-2

Chemical Structure
2.16.840.1.113883.3.26.1.1
C103240

ED
application/x-inchi-key
http://www.w3.org/2001/XMLSchema-instance
OBBXFSIWZVFYJR-UHFFFAOYSA-L
```
Example: racemic mixture

<table>
<thead>
<tr>
<th>name</th>
<th>METHYLEPHEDRINE, (+/-)-</th>
</tr>
</thead>
<tbody>
<tr>
<td>mixture component</td>
<td>2.16.840.1.113883.3.26.1.1</td>
</tr>
<tr>
<td>C103243</td>
<td></td>
</tr>
</tbody>
</table>
Example: chemical mixture with undefined ratio of components

- Name: SENNOSIDES A AND B
- Mixture component:
  - Code System: 2.16.840.1.113883.3.26.1.1
  - Code: C103243
  - URL: http://www.w3.org/2001/XMLSchema-instance
  - URG_PQ

- Numerator:
  - Value: 0
  - Unit: 1
  - Inclusive: false

- Denominator:
  - Value: 1
  - Unit: 1
Elements of Substance Definition (continuation)

- **Moiety** may be defined by references
  - **Moiety** may reference another substance (defined in another index file)
  - **Moiety** may reference *universal structural unit* defined in the same index file

- **Moiety** may have one or more **bonds** that indicate how it connects to other moieties
  - **Bond** has a type
  - **Bond** has position on distal moiety
Example: Antibody-drug conjugate

**Universal structural units**

**Heavy chain**
QIQLQSGPEVKPGASVKSQCKASGTYTFTDYITWVKQPGQGLEWIGWIGWISGNTKYNFKKGA
TLTVDTSSSTAFMQLSLTEDTAVFCANYGNFYAYWQGQTQVTVSAASTKPSVLLFAPLAPSSTSTSG
GTAAAALGCVKDYFEPVPVTWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPPSVSSLTQYICNVNHK
PSNTKVDKVEPKSCDKTHCPCCPAELLGPGSLFPPKPKDTPDLTMSRTPEVTCVVVDVHEDPEVKG
NYVVDGVEVHNAKTPREEQYNSTYRVSVTLHVQLNLNGKYCKVSNKALPAPIEKTISKAKQG
PREPVYTLPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPVLDSDGFFLYSKVT
DSRWQQGNVFSCSVMHEALHNHYTQKSSLSPG

**Light chain**
DIVLTQSPASLAVSLGQRATISCKASQSVDFDGYMNWYQQPKPGQPVKWYASSNLGSIGPARFSGS
GSGTDFTLNIHPVEEEDAATYQCQSNEDPWTGGGKLEIKRTVAAPSVFIFPPSDEQLKSGTSVVC
LLNNFYPREAVKQVVDNALSQNSSEQVTEQDSDKSTSYSLSTLSSKADYEHKVAYACVYITEQGLS
SPVTKSFRGEC

**Particular moieties**

**drug**

brentuximab = Ig(3H)_n, n = 3 - 5
Example: heavy chain of antibody

<table>
<thead>
<tr>
<th>identifiedSubstance</th>
<th>HCh</th>
</tr>
</thead>
<tbody>
<tr>
<td>code</td>
<td>b2315501-dc0c-56d0-e044-001185133a64</td>
</tr>
<tr>
<td>name</td>
<td>IgG Heavy Chain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>characteristic</th>
<th>X88850-0</th>
</tr>
</thead>
<tbody>
<tr>
<td>code</td>
<td>amino acid sequence, complete</td>
</tr>
<tr>
<td>codeSystem</td>
<td>2.16.840.1.113883.3.26.1.1</td>
</tr>
<tr>
<td>value</td>
<td>ED</td>
</tr>
<tr>
<td>xsi:type</td>
<td>application/x-aa-seq</td>
</tr>
<tr>
<td>mediaType</td>
<td>QIQLQQSGPEVVKPGASVKISCKASGYTFTDDYTITWVKQKPGQGLEWIGWIYPGSGNTKYNFKGK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>moiety</th>
<th>C8882-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>code</td>
<td>2.16.840.1.113883.3.26.1.1</td>
</tr>
<tr>
<td>displayName</td>
<td>protein sub-unit</td>
</tr>
<tr>
<td>quantity</td>
<td></td>
</tr>
<tr>
<td>numerator</td>
<td></td>
</tr>
<tr>
<td>value</td>
<td>1</td>
</tr>
<tr>
<td>unit</td>
<td>mol</td>
</tr>
<tr>
<td>denominator</td>
<td></td>
</tr>
<tr>
<td>value</td>
<td>1</td>
</tr>
<tr>
<td>unit</td>
<td>mol</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>partMoiety</th>
<th>SU1</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>b1f70df7-8153-45c1-9a21-6e1a1b4c97e</td>
</tr>
<tr>
<td>extension</td>
<td>HCh</td>
</tr>
<tr>
<td>root</td>
<td>b2315501-dc0c-56d0-e044-001185133a64</td>
</tr>
<tr>
<td>code</td>
<td>IgG Heavy Chain</td>
</tr>
<tr>
<td>codeSystem</td>
<td></td>
</tr>
<tr>
<td>name</td>
<td></td>
</tr>
</tbody>
</table>
Example: drug molecule connected to cysteine
Elements of Substance Definition (continuation)

• Substance may have non-structural characteristics
  • Chemical substance may have optical activity
  • Biological substance has an organism identifier
    ✓ genus
    ✓ species
    ✓ author
• Substance may have modification process
• etc.
FDA will release Substance Index Files in stages

What substances do already have Index Files?

- Public Domain Substances
- Listed as Ingredients in drug labels
- Chemical Substances

What substances will have Index Files?

- Public Domain Substances
- That have UNII
- Chemicals, proteins, polymers, nucleic acids, biological substances (e.g., vaccines)
Acknowledgements

Gunther Shadow (SPL)
Randy Levin (FDA)
Peggy Leizear (FDA)
Lonnie Smith (FDA)
Bill Hess (FDA)
Eva Tu (FDA)
Jim Shugars (FDA)
Vada Perkins (FDA)
Dammika Amugoda-Kankanange (FDA)
Iwona Weidlich (FDA)
Archana Newatia (FDA)

Andrzej Wilk (USP)
Valery Tkachenko (RSC)
Igor Filippov (NIH)
Dmitry Ivanov (NIH)
Marc Nicklaus (NIH)
Evan Bolton (NIH)
Noel Southall (NIH)
Tyler Peryea (NIH)
Wolf Ihlenfeldt (Xemistry)
Anne Kling (Mitre)