SPL Index Files for substances: means for data exchange and substance identification

Yulia Borodina, Larry Callahan, Frank Switzer, Randy Levin, Gunther Shadow*, Valery Tkachenko**

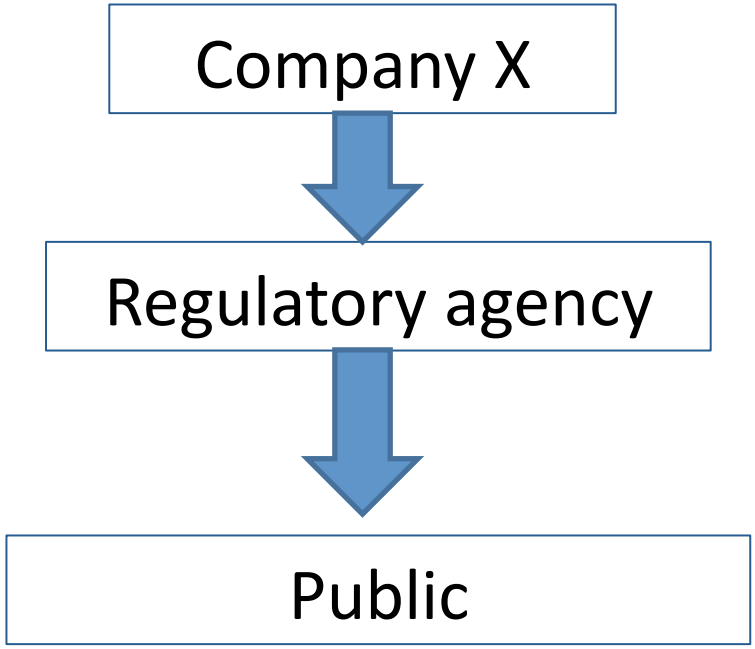
FDA/OIM/Medical Informatics Staff , *Pragmatic Data LCC, **RSC

What is SPL?

The Structured Product Labeling (SPL) is a document markup standard approved by Health Level Seven (HL7) and adopted by FDA as a mechanism for exchanging product and facility information.

The HL7 SPL was officially agreed to in the ICH meeting in Fukuoka (June 2012).

SPL Exchange



What is the SPL Substance Index File?

An implementation of IDMP ISO 11238 standard

- IDMP Implementation Guide explicitly references HL7 SPL standard

A mechanism for exchanging substance information

A basis for system-independent substance identification

Factors that determine structure of SPL Substances Index File

- HL7 v3 standard (health informatics)
- IDMP ISO 11238 standard (health informatics)
- Data of the FDA Substance Registration System (SRS)
- Standards used in Chemo-Bio informatics (e.g. InChI)

Key principles of SPL Substance Index File

- Every identifier (except InChI and InChI-key) must have OID
- Every concept must be associated with publicly recognized code system and have code and code system identifier
- Every chemical structure must be InChI-able without loss of information. InChI is the main chemical identifier.
- Every SPL Substance Index File should have all information necessary for substance identification (self sufficient message)

Every identifier must have OID

Example: Unique Ingredient Identifier

UNII

OID for UNII

<code code="FY36J270ES" codeSystem="2.16.840.1.113883.4.9" />

Every concept must have code and code system identifier

Example: "Chemical structure" concept

Code for concept "Chemical Structure"

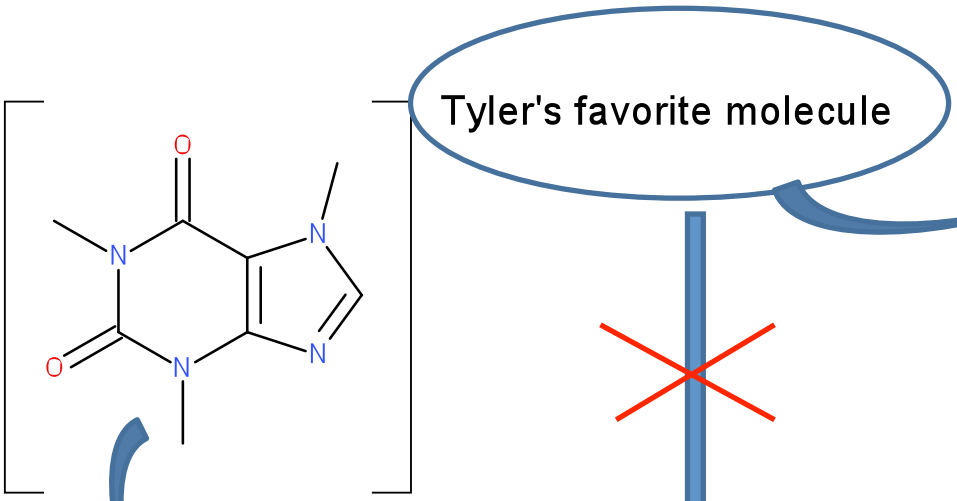
```
<characteristic>
    <code displayName="Chemical Structure"
codeSystem="2.16.840.1.113883.3.26.1.1" code="C103240" />
```

OID for "NCI Thesaurus" code system

Every chemical structure must be InChI-able without loss of information

Example: S-Text in Molfiles

Some other part of SPL XML

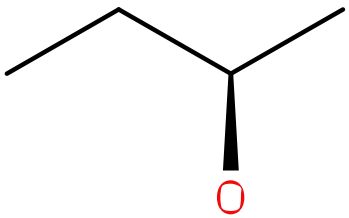


InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

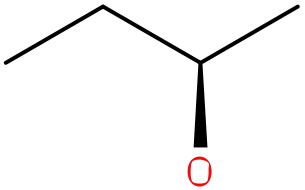
Structure part of SPL XML

Every chemical structure must be InChI-able without loss of information

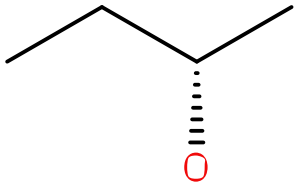
Example: Enhanced stereo



AND Enantiomer



InChI=1S/C4H10O/c1-3-4(2)5/h4-5H,3H2,1-2H3/t4-/m1/s1

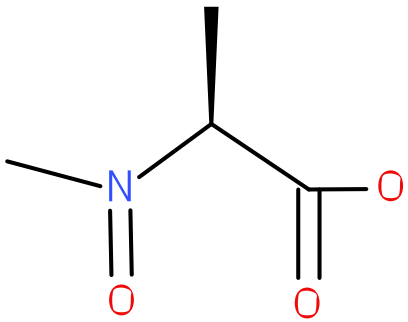


InChI=1S/C4H10O/c1-3-4(2)5/h4-5H,3H2,1-2H3/t4-/m0/s1

SPL file can carry multiple Molfiles/ InChIs

Every SPL Substance Index file should have all information necessary for substance identification (self sufficient message)

Example: System X has its own controlled vocabularies for modified amino acids



N-Acetyl-D-alanine

NA-dA

SPL file

Chemical structure representation in SPL file


```
<characteristic>
  <code displayName="Chemical Structure" codeSystem="2.16.840.1.113883.3.26.1.1" code="C103240" />
  <value p4:type="ED" mediaType="application/x-mdl-molfile" xmlns:p4="http://www.w3.org/2001/XMLSchema-instance">
 -INDIGO-01241311212D

45 45  0  0  0  0  0  0  0  0999 V2000
  13.7542  -12.5750    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  12.6292  -11.9417    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  13.7542  -13.8375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0


 <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" xmlns:p4="http://www.w3.org/2001/XMLSchema-instance">
 InChI=1S/C39H66O6/c1-4-7-10-13-16-19-22-25-30-43-37(40)34-28-29-35(38(41)44-31-26-23-20-17-14-11-8-5-2)36(33-34)
 39(42)45-32-27-24-21-18-15-12-9-6-3/h28-29,33H,4-27,30-32H2,1-3H3</value>
 </characteristic>


  <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-key" xmlns:p4="http://www.w3.org/2001/XMLSchema-instance">
  SMYKBXMWXCZOLU-UHFFFAOYSA-N</value>
  </characteristic>
```

Structural representation type is specified as a MIME media type code

Type	MIME Media Type
InCHI	application/x-inchi
MOLFILE	application/x-mdl-molfile
SMILES	application/x-smiles
CHIME	application/x-mdl-chime
Amino Acid Letter Sequence	application/x-aa-seq
DNA Sequence	application/x-dna-seq
RNA Sequence	application/x-rna-seq

“Stereochemistry” element

Stereochemistry must be indicated on the structure

Special stereo such as square planar, tetrahedral and octahedral must be indicated in the XML

```
<characteristic>
 <code code="C18188" codeSystem="2.16.840.1.113883.3.26.1.1" displayName="Stereochemistry Type" />
  <value xsi:type="CV" code="C103215" codeSystem="2.16.840.1.113883.3.26.1.1“ displayName="TETRAHEDRAL" />
</characteristic>
```

“Optical activity” element

```
<characteristic>
  <code code="C103201" codeSystem="2.16.840.1.113883.3.26.1.1" displayName="Optical Activity" />
  <value xsi:type="CV" code="C103204" codeSystem="2.16.840.1.113883.3.26.1.1" displayName="NONROTATORY" />
</characteristic>
```

“Moiety” element

<moiety>
<code displayName="**mixture component**" codeSystem="**2.16.840.1.113883.3.26.1.1**" code="**C103243**"/>
…
<characteristic>
<code displayName="**Chemical Structure**" codeSystem="**2.16.840.1.113883.3.26.1.1**" code="**C103240**"/>
<value xsi:type="ED" mediaType="application/x-inchi">InChI=1S/C22H32O4/c1-3-4-7-15(2)20(23)11-10-18-19-13-16(8-5-6-9-22(25)26)12-17(19)14-21(18)24/h8,10-11,15,17-21,23-24H,5-7,9,12-14H2,1-2H3,(H,25,26)/b11-10+,16-8+/t15-,17+,18-,19+,20-,21-/m1/s1</value>
</characteristic>
…
</moiety>
<moiety>
<code displayName="**mixture component**" codeSystem="**2.16.840.1.113883.3.26.1.1**" code="**C103243**"/>
…
<characteristic>
<code displayName="Chemical Structure" codeSystem="2.16.840.1.113883.3.26.1.1" code="C103240"/>
<value xsi:type="ED" mediaType="application/x-inchi">InChI=1S/C22H32O4/c1-3-4-7-15(2)20(23)11-10-18-19-13-16(8-5-6-9-22(25)26)12-17(19)14-21(18)24/h8,10-11,15,17-21,23-24H,5-7,9,12-14H2,1-2H3,(H,25,26)/b11-10+,16-8+/t15-,17-,18+,19-,20+,21+/m0/s1</value>
</characteristic>
…
</moiety>

“Quantity” element

<moiety>
<code displayName=“**chemical moiety**" codeSystem="**2.16.840.1.113883.3.26.1.1**" code="**C103246**"/>
<quantity>
<numerator xsi:type="URG_PQ" value="">
<low value="0.5" unit="1"/>
<high value="1" unit="1"/>
</numerator>
<denominator value="1" unit="1"/>
</quantity>
…
<characteristic>
<code displayName="**Chemical Structure**" codeSystem="**2.16.840.1.113883.3.26.1.1**" code="**C103240**"/>
<value xsi:type="ED" mediaType="application/x-inchi">InChI=1S/C22H32O4/c1-3-4-7-15(2)20(23)11-10-18-19-13-16(8-5-6-9-22(25)26)12-17(19)14-21(18)24/h8,10-11,15,17-21,23-24H,5-7,9,12-14H2,1-2H3,(H,25,26)/b11-10+,16-8+/t15-,17+,18-,19+,20-,21-/m1/s1</value>
</characteristic>
…
</moiety>

Substance identification based on SPL Substance Index file

<code code="**567ff9e0dcea302d86c6be47591c3e83**" codeSystem="**2.16.840.1.113883.3.2705**"/>

Hash code computed based on InChI codes for moieties and sub-moieties

SPL Hash code vs UNII code

UNII is generated by Substance Registration System

Hash code is generated by external application that analyses SPL XML exported from SRS

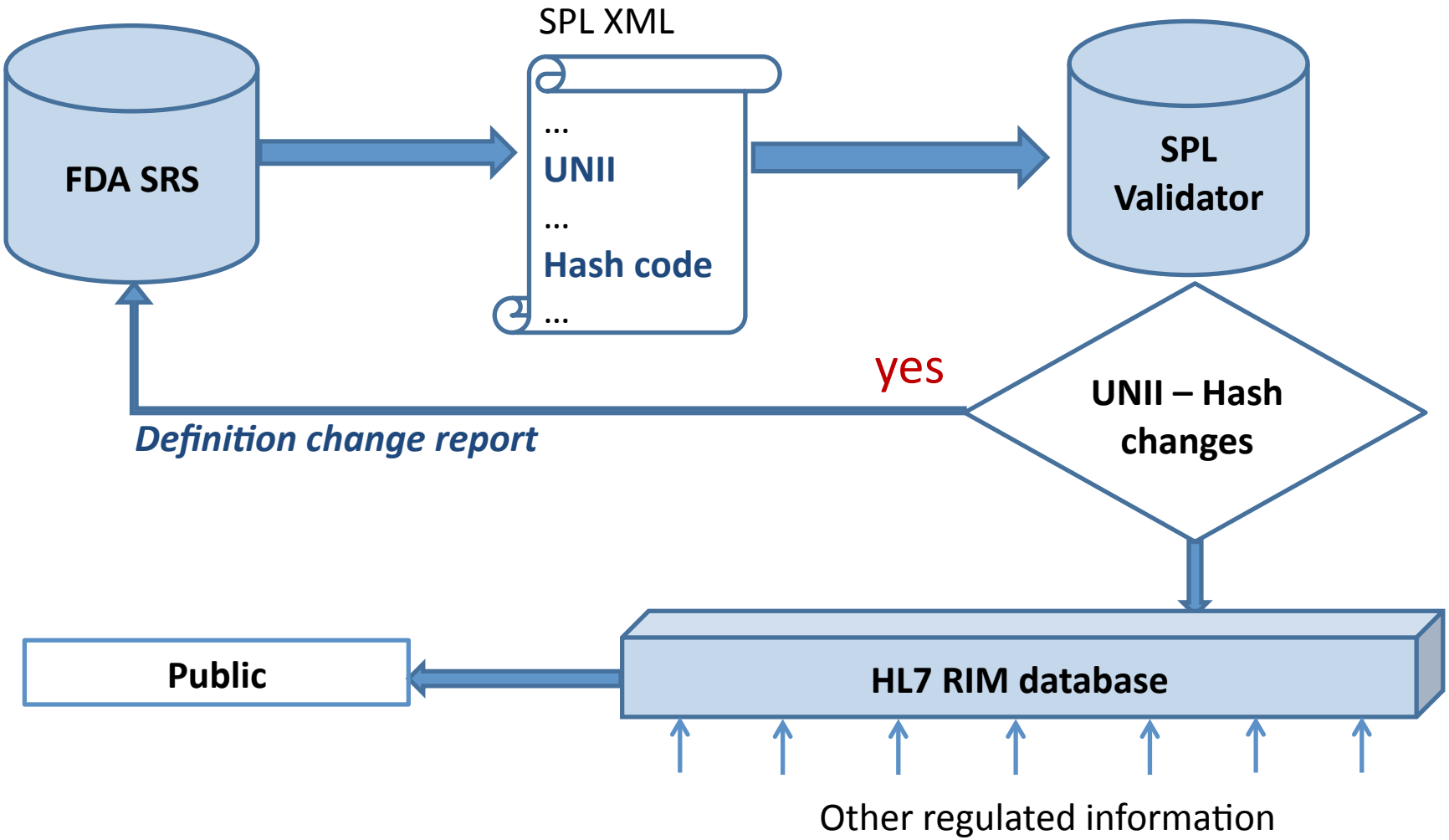
UNII is randomly generated identifier

Hash code is function of data

UNII code is assigned under human control

Hash code is assigned automatically

Using hash code for validating substance definitions



Current state of the project

- SPL Substance Index files for chemical substances/ chemical mixtures is in the test stage
  - ~5000 SRS substances that are in marketed drugs are converted into SPL XMLs
  - After internal validation SPL files will be published on Daily Med
- SPL Substance Index files for proteins is in the development stage
- SPL Substance Index files for polymers is in the data analysis stage
- SPL Substance Index files for structurally diverse substances needs clarification of the concept “organism”

Advantages of using SPL Substance Index Files

- Integration of chemoinformatics with health informatics
- Easy data exchange between regulatory agencies
- Industry can build stand-alone applications/systems
- System-independent identification of substances

Support of SPL Substance Index Files in other systems

- CACTVS toolkit will provide SPL reader starting from version 3.410
- PubChem is interested in supporting SPL XMLs

Acknowledgments

- Peggy Liezear (FDA)
- Eva Tu (FDA)
- Markus Sitzmann (NCI)
- Wolf Ihlendfeldt (Xemistry GmbH)