Harmonizing EHR Databases with SNOMED CT

Nor Azlinayati Abdul Manaf
Dr. Khalil Bouzekri
Dr. Dickson Lukose

SNOMED CT Implementation Showcase 2014, Amsterdam
30th October 2014
Outline

• Challenges of Data Harmonization in Malaysia
• Mi-Harmony Tool
• Result and Findings
• Summary & Conclusion
Data Harmonization Process

Freetext Challenge

Role: Consolidate the data
- Time Consuming
- Limited Resources

Role: Interpret the data
- Automate processes
- Enable Semantic
- Most Knowledgeable
- Multilingual

Subject Matter Expert

IT Staff

© 2014 MIMOS Berhad
## Challenge of freetext

### HIS1

<table>
<thead>
<tr>
<th>ID</th>
<th>Patient Name</th>
<th>Gender</th>
<th>Symptoms</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>xx</td>
<td>M</td>
<td>Fatigue, Chest pain</td>
<td>Heart failure</td>
</tr>
<tr>
<td>2</td>
<td>yy</td>
<td>F</td>
<td>Breathlessness</td>
<td>Acute asthma</td>
</tr>
</tbody>
</table>

### HIS2

<table>
<thead>
<tr>
<th>ID</th>
<th>Nama Pesakit</th>
<th>Jantina</th>
<th>Tanda-tanda</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>aa</td>
<td>L</td>
<td>Letih, Sakit dada</td>
<td>Lemah jantung</td>
</tr>
<tr>
<td>2</td>
<td>bb</td>
<td>P</td>
<td>Sesak nafas</td>
<td>Lelah</td>
</tr>
</tbody>
</table>

### HIS3

<table>
<thead>
<tr>
<th>ID</th>
<th>Pat.Name</th>
<th>Sex</th>
<th>Disorder</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kk</td>
<td>Male</td>
<td>Tiredness, Pain in chest</td>
<td>Cardiac failure</td>
</tr>
<tr>
<td>2</td>
<td>ll</td>
<td>Female</td>
<td>Dyspnea</td>
<td>Acute bronchitis</td>
</tr>
</tbody>
</table>
Outline

• Challenges of Data Harmonization in Malaysia

• Mi-Harmony Tool

• Result and Findings

• Summary & Conclusion
SNOMED CT based Clinical Data Harmonization

Existing HIS

ETL

Data Aggregation

Un-coded Databases

Data Consolidation

Coded Databases

Information Visualization

SNOMED CT RefSets

Cube Generation

MYHDW
SNOMED CT based Clinical Data Harmonization

- Semi-automated
- Auto-suggestion of SNOMED CT codes
SNOMED CT based Clinical Data Harmonization

1. Reference Sets
2. DB Schema
3. eHIS Database (DB)
4. eHIS-SNOMED Mapping
5. Mapping tables
6. Reverse Engineering (RE)
7. SNOMED Codification
8. SNOMED CT Coded Data

© 2014 MIMOS Berhad
Mi-Harmony

What is Mi-Harmony?
Learn More

© 2013 - MIMOS Berhad
Mi-Harmony: Mapping Table Columns
Mi-Harmony: Mapping Reference Table

### Database Table(s)
- AE_BAY_TYPE
- AE_DISASTER_TYPE
- AE_GCS_FOR_RESPONSE
- AE_GCS_RESPONSE
- AE_GCS_SCORE
- AE_IMMOBILIZATION
- AE_INJURY_CATEGORY
- AE_INJURY_SUBCATEGORY
- AE_PARAM

### Refset: Cardiology_Refset_1.1
- 243796009 Situation with explicit context
- 363787002 Observable entity
- 365580001 Finding of personal status
- 404684003 Clinical finding
- 71388002 Procedure

### Table Column(s)
- DISASTER_TYPE_CODE
- LONG_DESC
- SHORT_DESC
- EFF_STATUS
- ADDED_BY_ID
- ADDED_DATE
- ADDED_AT_WS_NO
- ADDED_FACILITY_ID
- MODIFIED_BY_ID

### Column Value(s)
- Air Crash
- Chemical Warfare
- Explosion
- Landslide
- Biomedical Warfare
- Collapsed Building
- Fire
- Maritime Accident
- Biological Warfare

### Mapping Concept(s)
<table>
<thead>
<tr>
<th>Selected value</th>
<th>Mapped Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Crash</td>
<td>500320005 (Air crash)</td>
</tr>
<tr>
<td>Fire</td>
<td>257195006 (Fire)</td>
</tr>
</tbody>
</table>

### Suggested Concept(s)
- 136953001 Fireman (87.14%)
- 423624000 Fireback (85.00%)
- 217177006 House fire (85.36%)
- 30348005 Fire (91.25%)
- 10727001 Fire salamander (85.36%)
- 257195008 Fire (100.00%)

© 2013 - MIMOS Berhad
Mi-Harmony: Codification
Mi-Harmony: Sample codified result

<table>
<thead>
<tr>
<th>COMPLAINT_ID</th>
<th>COMPLAINT_DESC</th>
<th>COMPLAINT_TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC03</td>
<td>Cardiac arrest (traumatic)</td>
<td>[cardiac arrest $397912004$, [unspecified $10003008$]]</td>
</tr>
<tr>
<td>CC04</td>
<td>Chest pain</td>
<td>[chest pain $139228007$, [unspecified $10003008$]]</td>
</tr>
<tr>
<td>CC08</td>
<td>Edema, generalized</td>
<td>[Generalized edema $442433009$]</td>
</tr>
<tr>
<td>CC11</td>
<td>Unilateral reddened hot limb: DVT symptoms</td>
<td>[other specified soft tissue $181789000$, [disorders $64572001$, unspecifed site $64556001$]]</td>
</tr>
<tr>
<td>CC12</td>
<td>Earache</td>
<td>[Otalgia $155251005$]</td>
</tr>
<tr>
<td>CC13</td>
<td>Foreign body</td>
<td>[Foreign body in ear $75441006$]</td>
</tr>
<tr>
<td>CC14</td>
<td>Loss of hearing</td>
<td>[hearing loss $155242002$, [unspecified $10003008$]]</td>
</tr>
<tr>
<td>CC15</td>
<td>Tinnitus / Dysacusis</td>
<td>[Tinnitus $60862001$]</td>
</tr>
<tr>
<td>CC16</td>
<td>Discharge</td>
<td>[Otorrhoea $65666001$]</td>
</tr>
<tr>
<td>CC17</td>
<td>Ear injury</td>
<td>[Superficial injury of the ear $1910005$]</td>
</tr>
<tr>
<td>CC18</td>
<td>Dental / Gum problem</td>
<td>Disorder of gingiva [and $421829000$, [edentulous alveolar ridge $33446008$, ...]</td>
</tr>
<tr>
<td>CC19</td>
<td>Facial trauma</td>
<td>Superficial injury [of $18720000$, other parts [of $18720000$, the [head $69536$...</td>
</tr>
<tr>
<td>CC20</td>
<td>Sore throat</td>
<td>[acute pharyngitis $155502006$, [unspecified $10003008$]]</td>
</tr>
<tr>
<td>CC21</td>
<td>Neck swelling/pain</td>
<td>Localized [swelling $65124004$, mass and [lump $300848003$, [neck $30255$, ...</td>
</tr>
<tr>
<td>CC23</td>
<td>Difficulty swallowing / Dysphagia</td>
<td>Other unspecified dysphasia</td>
</tr>
<tr>
<td>CC25</td>
<td>Epistaxis</td>
<td>[Epistaxis $12441001$]</td>
</tr>
<tr>
<td>CC27</td>
<td>Foreign body, nose</td>
<td>[Foreign body in nostril $33890007$]</td>
</tr>
<tr>
<td>CC28</td>
<td>URTI complaints</td>
<td>[acute upper respiratory infection $155516001$, [unspecified $10003008$]]</td>
</tr>
</tbody>
</table>
Outline

• Challenges of Data Harmonisation in Malaysia
• Mi-Harmony Tool
• Result and Findings
• Summary & Conclusion
Evaluation & Validation: Methodology

1. Prepare Gold Standard:
   - Identified 100 terms/phrases from HIS database table
     • Chief complaints
   - Manually tagged the relevant SNOMED CT codes for the terms/phrases
     • 3 medical officers were tasked to perform this task

2. Compare with Gold Standard:
   - Compare the codes generated by Mi-Harmony tool with the one produced in Gold Standard
Evaluation & Validation: Findings

• Precision: 90%
  – 10% of the terms/phrases are coded incorrectly
    • Concept selected from incorrect hierarchy
      – Eg: *Superficial injury of unspecified body region* was codified with 243928005 which refers to * Entire body region (body structure)*
    • Concept selected more specific rather than generic
      – Eg: *Cyanosis* was codified with 119419001 which refers to *Cyanosis of skin (finding)* instead of 3415004 *Cyanosis (finding)*

• Recall: 70%
  – 30% of the terms/phrases are not coded even though the concept exists in SNOMED CT
    • Use of abbreviations and acronyms in the medical text
    • No exact matching of the phrases in SNOMED CT
Mi-Harmony: Ongoing Enhancement

• Improving the concept matching algorithm
  – Include context-based concept matching
    • To solve the problem of concept selection from incorrect hierarchy

• Improving and enhancing the list of commonly used acronyms, abbreviations and local terms that can be found in HIS databases
  – To improve natural language understanding of the free-text in medical records
Mi-Harmony: Ongoing Enhancement

• Ref Set authoring activity – Cardiology
• Identify the relevant SNOMED CT semantic category for the Patient Registry form

Summary & Conclusion

• Current challenges in data harmonisation for HIS in Malaysia

• Mi-Harmony as a tool for data harmonisation utilizing SNOMED CT and data from HIS databases

• Further enhancement and validation required to improve the accuracy of the tool
Contact Us:
Nor Azlinayati Abdul Manaf
Head Of Knowledge Engineering Team,
MIMOS Berhad, Malaysia
Azlinayati.Manaf@mimos.My
Enabling the future

Freetext Challenge

Today

Subject Matter Expert

- Time Consuming
- Limited Resources
- Skills-dependent
- Prone to errors

Inaccurate Information

Tomorrow

Mi-Harmony

- Automate processes
- Deal with freetext
- Standardize content

Accurate Information

© 2014 MIMOS Berhad