MINISTRY OF HEALTH
MALAYSIA

Performing analytics on SNOMED CT coded database, Serdang Hospital use-case

SNOMED CT EXPO 2015

Montevideo, 29 October 2015

Dr. Khalil Bouzekri, Dr. Md Khadzir Sheikh Ahmad, Wael Hamdan, Dr. Dickson Lukose, Dr Nur Shaema Darus and Dr Syirahaniza Mohd Salleh

Agenda

MINISTRY OF HEALTH
MALAYSIA

1. **Start-up project: overall objective**

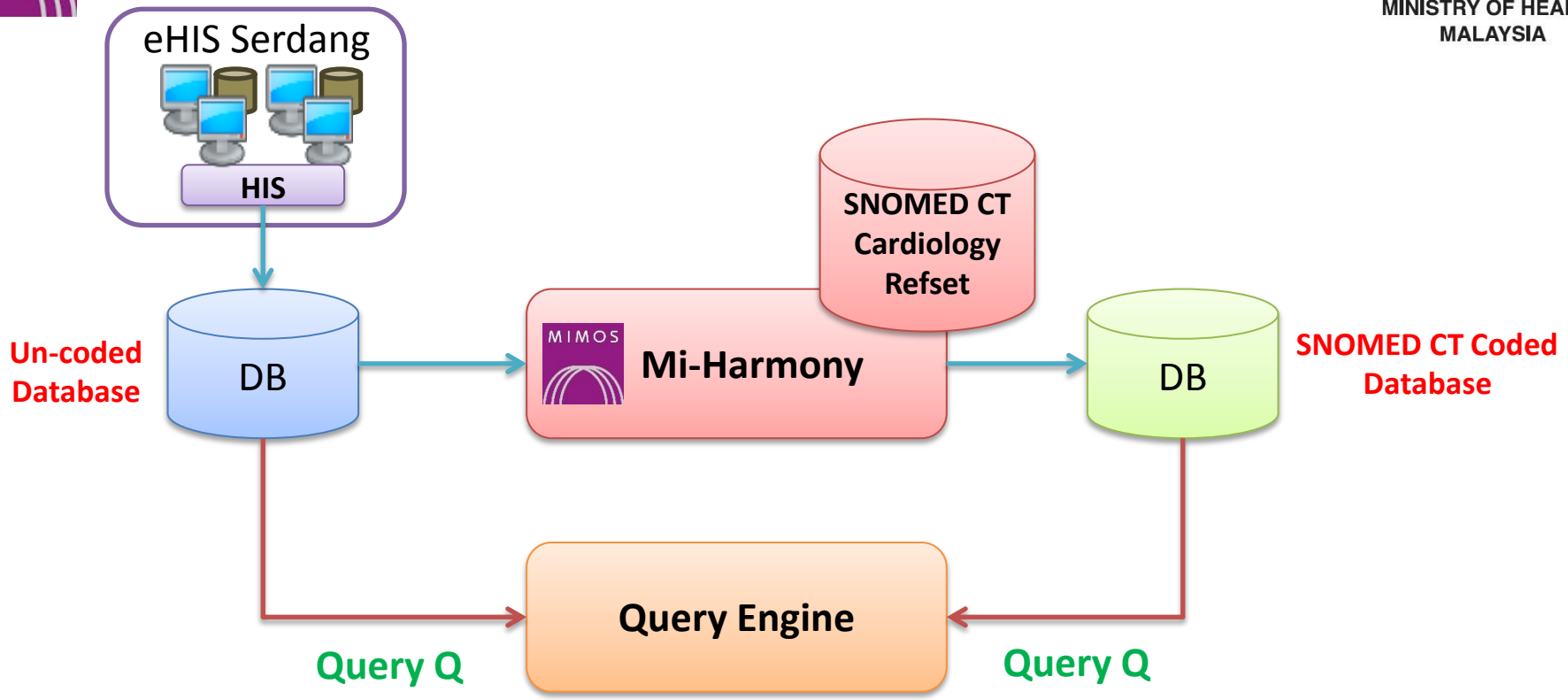
2. **Methodology to get SNOMED CT codified data**
   a. Prepare the data from eHIS database
   b. Codify the data with SNOMED CT
   c. Validate the codification

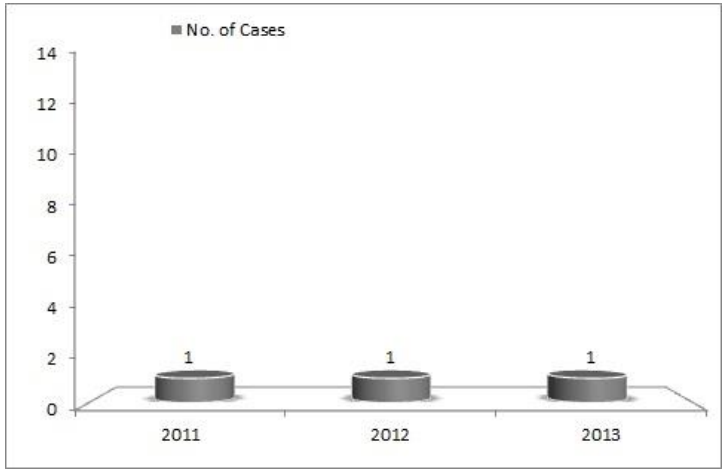
3. **Analytics over SNOMED CT codified data**

4. **Challenges and Future work**
   a. SNOMED CT versioning
   b. Performance
   c. Integration into the Malaysian Health Data Warehouse

Start-up project: overall objective

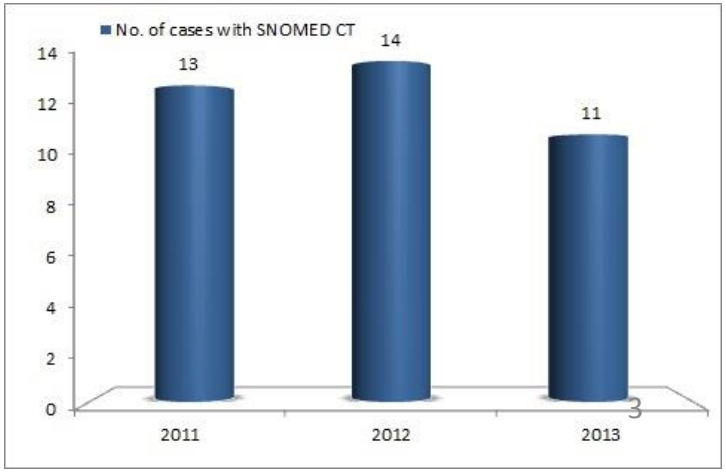
MINISTRY OF HEALTH MALAYSIA





Same Query

Different Results?



Agenda

MINISTRY OF HEALTH MALAYSIA

1. **Start-up project: overall objective**

2. **Methodology to get SNOMED CT codified data**
   a. Prepare the data from eHIS database
   b. Codify the data with SNOMED CT
   c. Validate the codification

3. **Analytics over SNOMED CT codified data**

4. **Challenges and Future work**
   a. SNOMED CT versioning
   b. Performance
   c. Integration into the Malaysian Health Data Warehouse

Prepare the data from eHIS database

MINISTRY OF HEALTH MALAYSIA

- eHIS Database:
  - 2K+ tables
  - Contain data from Cardiology Department

- Identify table with doctor-written patient diagnoses
  - PR_PROBLEM_ASSESSMENT (col. PROBLEM_REMARKS)
  - 491,612 records

- Identify patterns related to cardiology terms
  - 31 patterns identified: %heart%, %card%, %coronar%, %fibril%, %mitral%, %angio%, %angina%, %stemi%, %lad%, %2vd%, %3vd%, %ccs%, etc.

Prepare the data from eHIS database

MINISTRY OF HEALTH MALAYSIA

- Extract the data using identified cardiology related patterns: 10,692 records were extracted

- Filter cardiology related data: 10,228 records verified as Cardiology related diagnoses. Confirmation done by 2 verifiers and with crosschecking method

Examples of diagnosis:

- Missed anterior MI (killpi 3) , IHD h/o NSTEMI
- 3VD, Unstable angina (evolving ECG changes, Trop T negative), carotids normal
- NSTEMI and mixed with AF and atrial flutter

Codify the data with SNOMED CT

MINISTRY OF HEALTH MALAYSIA

- Use the Malaysian Cardiology Refset to guide the codification:
  - Built jointly by Cardiologists, Ministry of Health of Malaysia and MIMOS Berhad
  - Derived from the NCVD ACS & PCI (**N**ational **C**ardio**V**ascular disease **D**atabase patient registry form -**A**cute **C**oronary **S**yndrome & **P**ercutaneous **C**oronary **I**ntervention)
  - 380 existing SNOMED CT Concepts
  - 108 new concepts not existing in SNOMED CT
  - 30 post-coordinated concepts proposed
  - 1656 terms (including synonyms and short forms)

Codify the data with SNOMED CT

MINISTRY OF HEALTH MALAYSIA

- Examples of SNOMED CT codified diagnoses:

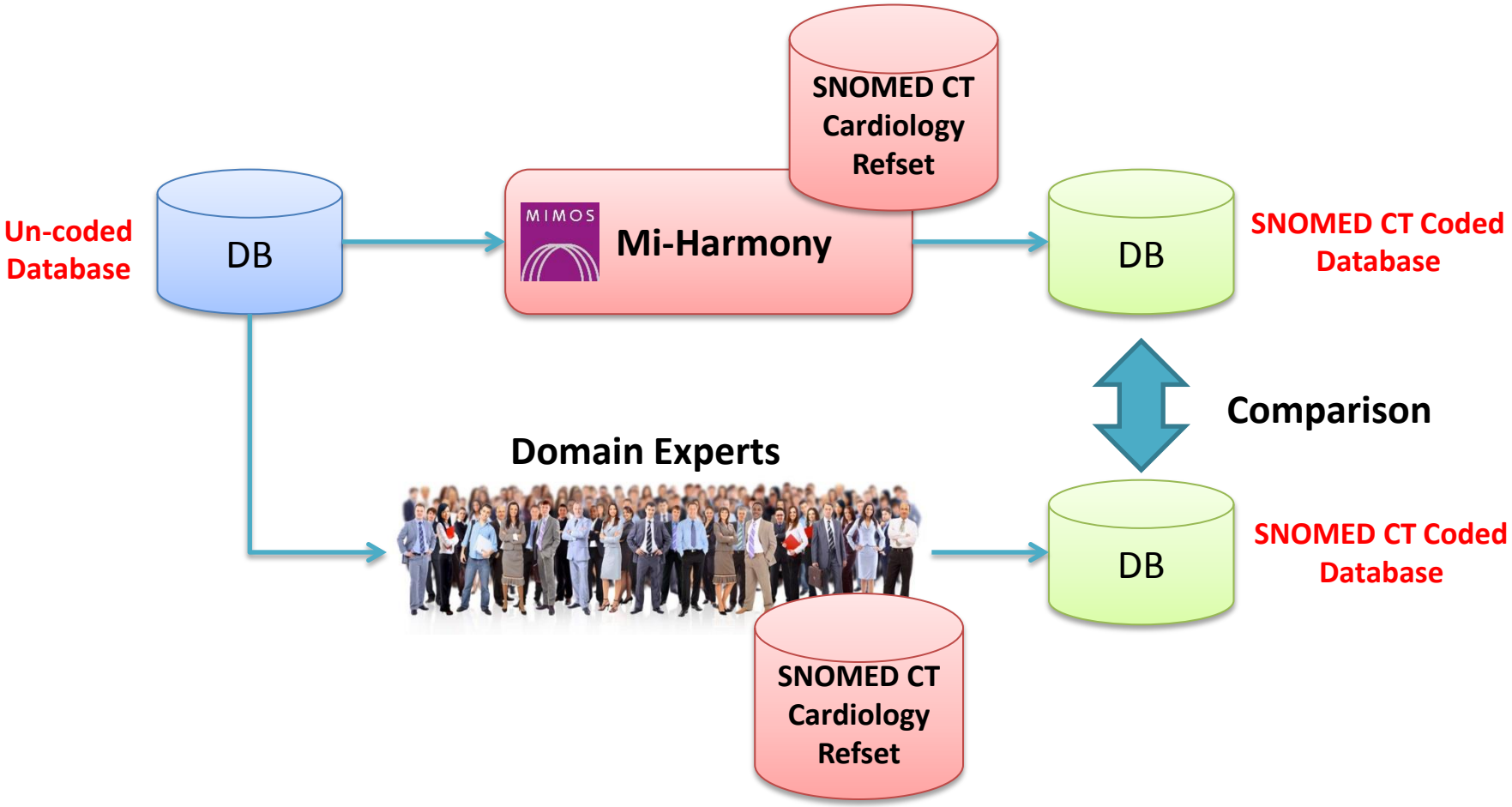
- Missed anterior MI ([killpi 3, #0000013#]) , [IHD, #414545008#] h/o [NSTEMI, #401314000#]

- 3VD, [Unstable angina, #4557003#] (evolving [ECG, #46825001#] changes , [Trop T, #102682001#] [negative, #260385009#]), carotids normal

- [NSTEMI, #401314000#] and mixed with [AF, #164889003#] and atrial flutter

Validating the codification

MINISTRY OF HEALTH MALAYSIA



Validating the codification

MINISTRY OF HEALTH MALAYSIA

- Results of the comparison over 100 diagnoses:
  - Precision: 95%
  - Recall: 94%

- For precision, we considered context-based terms to be wrongly codified, such as:
  - “symptoms of [heart disease, #56265001#]”
  - “h/o [NSTEMI, #401314000#]”

- For recall, the missed terms are mainly due to the informal way of writing. Example:
  - “CCS 1 - 2” which should be understood as “in between CCS 1 & CCS 2”

Agenda

MINISTRY OF HEALTH MALAYSIA

1. **Start-up project: overall objective**

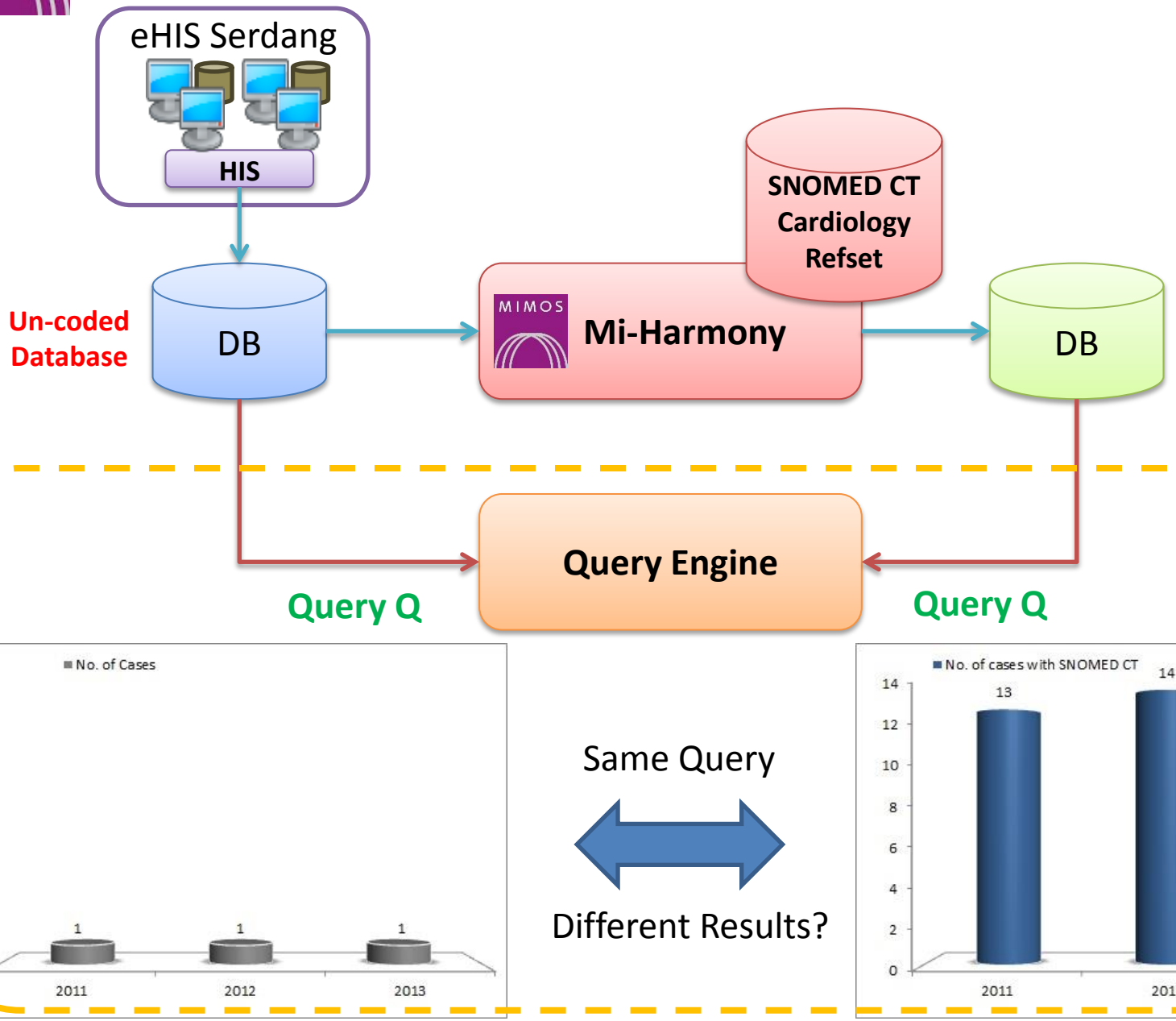
2. **Methodology to get SNOMED CT codified data**
   a. Prepare the data from eHIS database
   b. Codify the data with SNOMED CT
   c. Validate the codification

3. **Analytics over SNOMED CT codified data**

4. **Challenges and Future work**
   a. SNOMED CT versioning
   b. Performance
   c. Integration into the Malaysian Health Data Warehouse

Analytics over SNOMED CT codified data

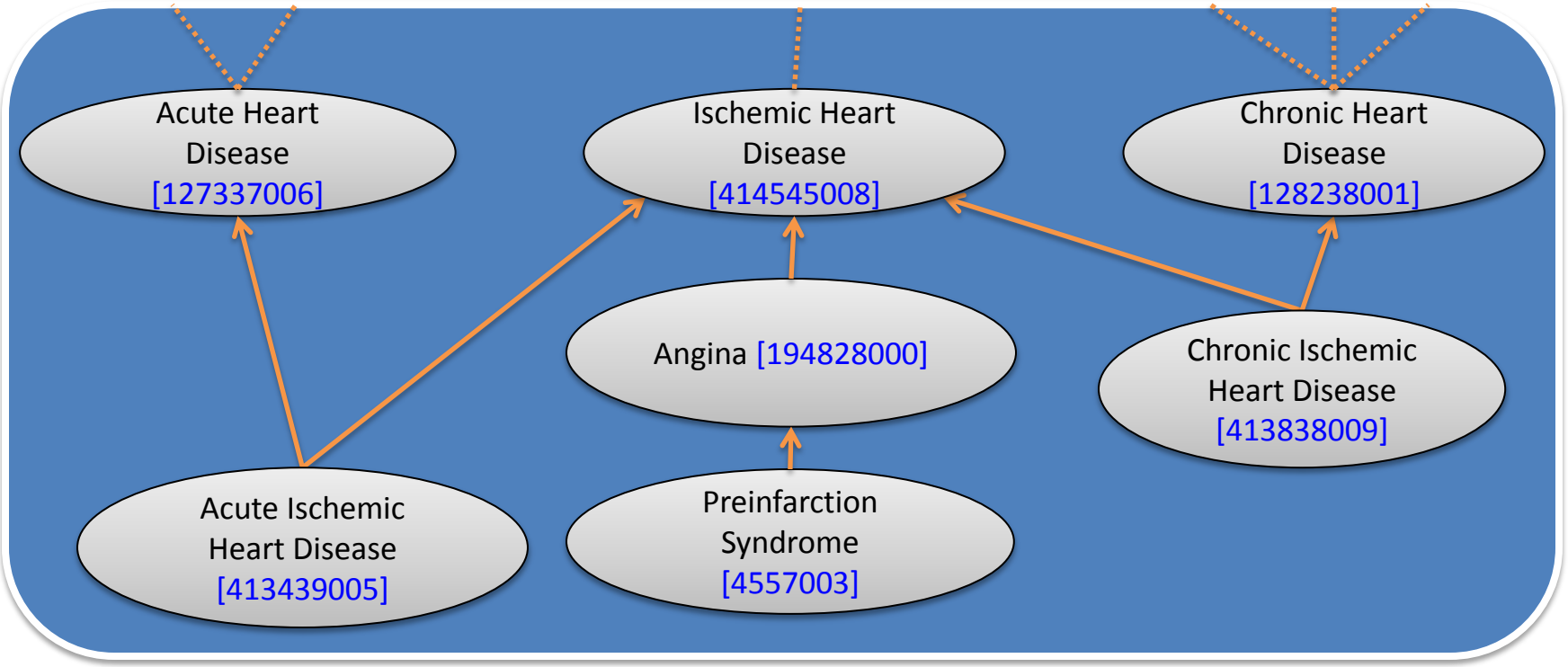
MINISTRY OF HEALTH MALAYSIA



Analytics over SNOMED CT codified data

MINISTRY OF HEALTH MALAYSIA

- Use of SNOMED CT at the reporting stage (fixed queries)
- Use of SNOMED CT subsumption relationships during query processing



Analytics over SNOMED CT codified data

MINISTRY OF HEALTH MALAYSIA

- Expand the query, guided by the Malaysian Cardiology Refset

- Example of input query:

- SELECT COUNT(*), YEAR(DATE) FROM PATIENT_DIAGNOSIS WHERE DIAGNOSIS LIKE "%#**Ischaemic Heart Disease**#%" GROUP BY YEAR(DATE);

- Example of expanded query:

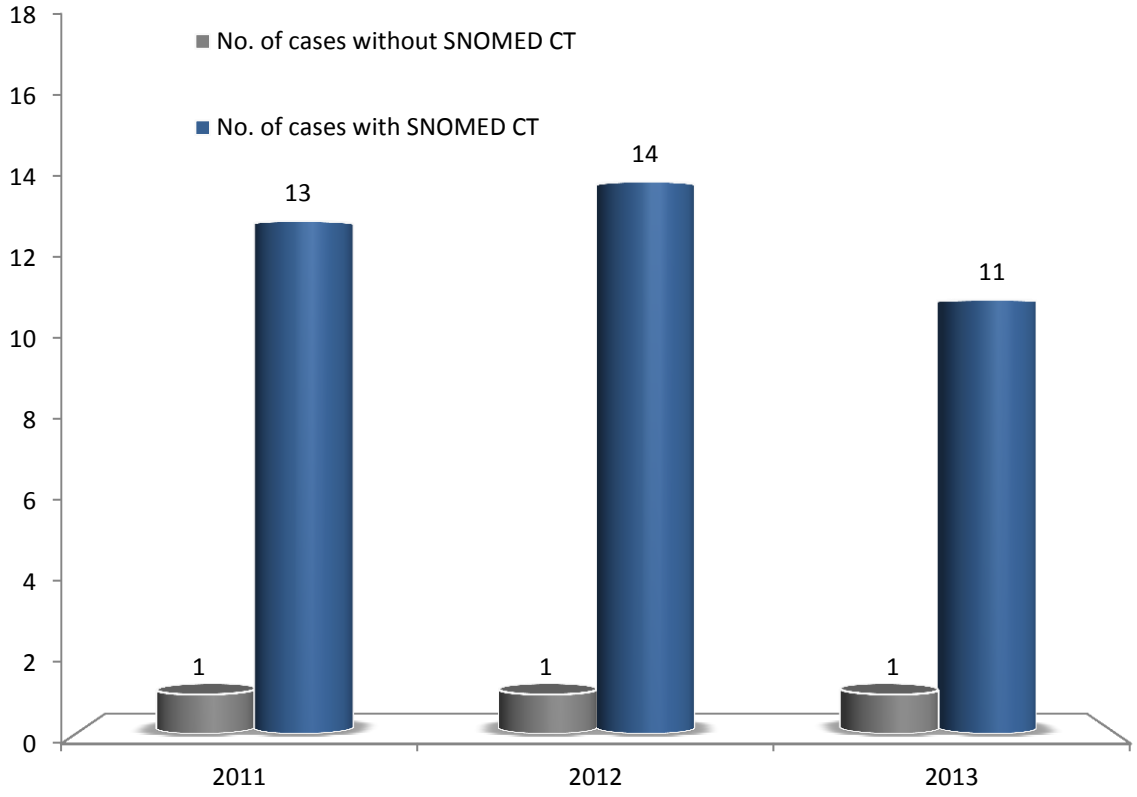
- SELECT COUNT(*), YEAR(DATE) FROM PATIENT_DIAGNOSIS WHERE DIAGNOSIS LIKE "%**#414545008#**%" OR DIAGNOSIS LIKE "%**#4557003#**%" OR DIAGNOSIS LIKE "%**#394659003#**%" GROUP BY YEAR(DATE);

"**414545008**" is the SNOMED CT code for "**Ischaemic Heart Disease**"
"**4557003**" is the SNOMED CT code for "**Unstable Angina**"
"**394659003**" is the SNOMED CT code for "**Acute Coronary Syndrome**"

Analytics over SNOMED CT codified data

MINISTRY OF HEALTH MALAYSIA

1. Results for 100 encounters manually validated



Agenda

MINISTRY OF HEALTH MALAYSIA

1. **Start-up project: overall objective**

2. **Methodology to get SNOMED CT codified data**
   a. Prepare the data from eHIS database
   b. Codify the data with SNOMED CT
   c. Validate the codification

3. **Analytics over SNOMED CT codified data**

4. **Challenges and Future work**
   a. SNOMED CT versioning
   b. Performance
   c. Integration into the Malaysian Health Data Warehouse

Handling SNOMED CT versioning

MINISTRY OF HEALTH MALAYSIA

- As per now:
  - Refset is derived from SNOMED CT INTERNATIONAL RELEASE
  - Coding is based on the Refset
  - Query expansion is using the same version of SNOMED CT INTERNATIONAL RELEASE

- What happens when we want to move to a newer SNOMED CT INTERNATIONAL RELEASE?
  - Refset has to be modified?
  - Behavior of query expansion has to be different?
    - Some data might have been codified with “old” codes while new data will use the “new” codes

Improving SNOMED CT Coding Performance

MINISTRY OF HEALTH
MALAYSIA

1. Improving Precision
    - Handling context of terms

2. Improving Recall
    - Increasing the number of handled short forms and acronyms
    - Handling certain forms of informal writing

3. Improving Speed
    - Moving to GPU-based parallel processing

Improving SNOMED CT Querying Performance

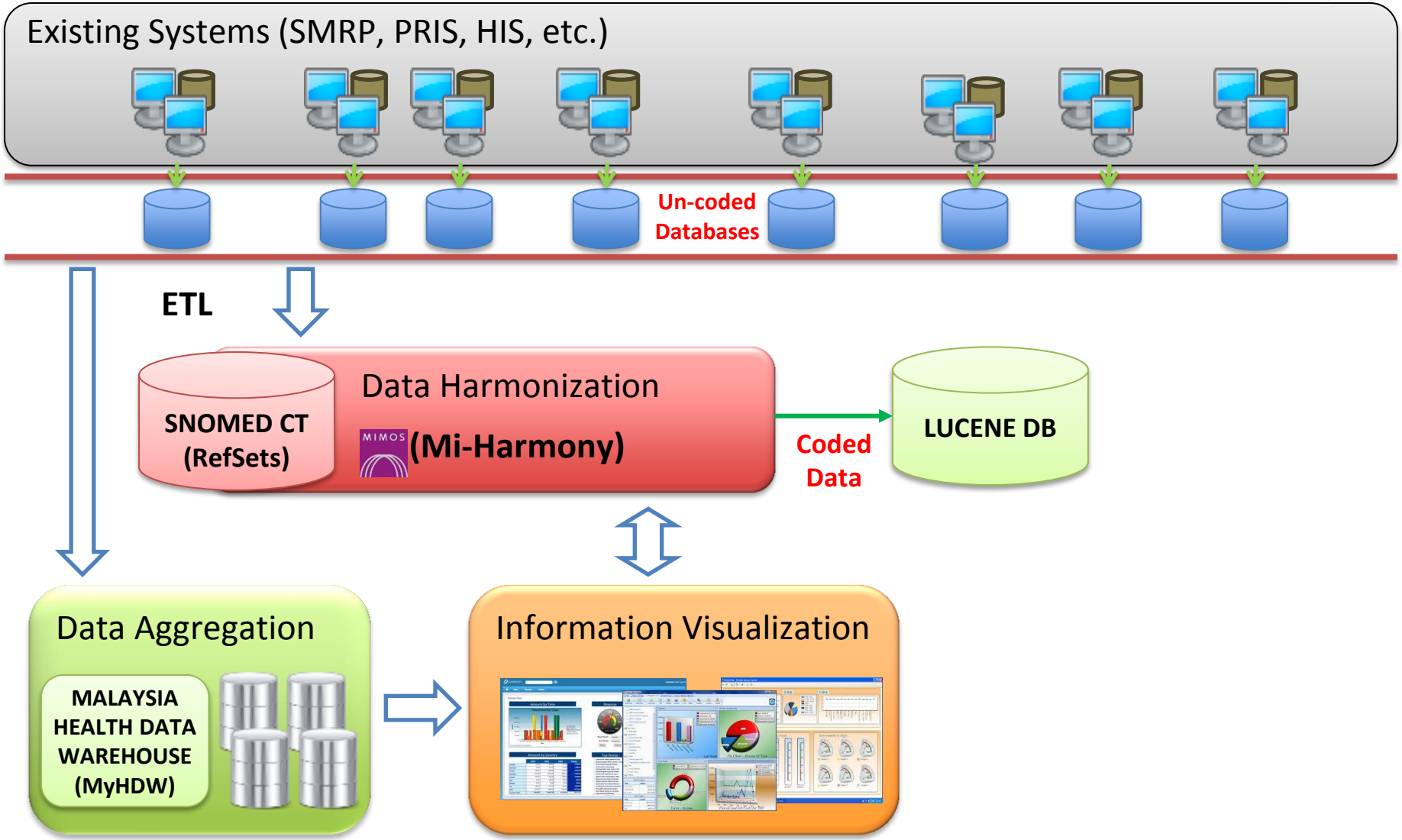
MINISTRY OF HEALTH
MALAYSIA

1. Improving Speed
   - Moving to GPU-based parallel processing

2. Handling versioning of SNOMED CT
   - To be investigated

Integration into MyHDW

MINISTRY OF HEALTH MALAYSIA



MIMOS

MINISTRY OF HEALTH
MALAYSIA

Q&A and Feedback

MINISTRY OF HEALTH
MALAYSIA

TERIMA KASIH

THANK YOU