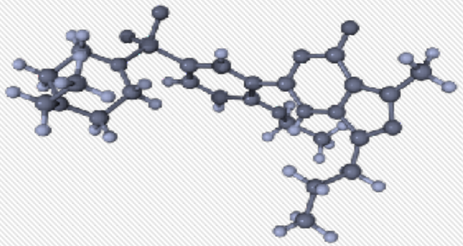
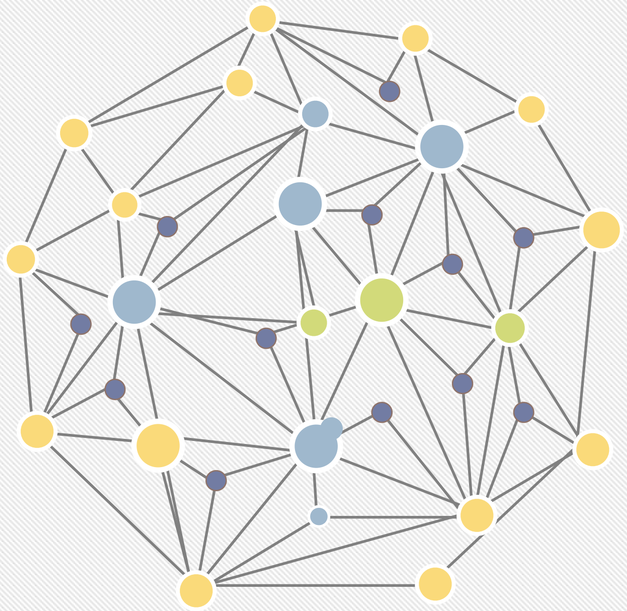
神州医疗
Digital China Health

# SNOMED CT Support Interoperability in Chinese Cancer Clinical Lab Test Data

**Zhi Wang, Zhouguang Hui, Qiang Guo, Yibo Gao, Jianyang Wang, Fang Wang, Mengchun Gong, Wenzhao Shi**
**Digital China Health**





PART 1

# Introduction of DCH

# Honor of DCH

## Partnership

- Fudan Pediatrics Hospital: Fudan Pediatrics—Digital China Health Research Center
- Broad Institute
- Philips


## Project

**2015: 863 Project《Data Analysis and Application for Malignant Tumor Big Data》**

**2016: National Key Research and Development Plan《Clinical Cohort Research for Rare Diseases》**

**2017: National Cancer Center and the platform**

**2018: NKRDP《The Research on the Construction of Clinical Big Data Platform and Biological Sample Library》**

**2018: NKRDP《New Service and Solution of Artificial Intelligence-based Clinical Decision Support》**

# China Health Medical Big Data Technology Development Group Corporation

National Health and Wellness Committee of the People's Republic

National Health Medical Big Data Security Management Committee

- China Health Medical Big Data Industry Development Group Corporation
- China Health Medical Big Data Technology Development Group Corporation
  - DCH
  - Other Company
- China Health Medical Big Data Co., Ltd.

**Health Care Big Data Industry Alliance**
**Vice President**

**Health Care Big Data Population Informationization Professional Committee**

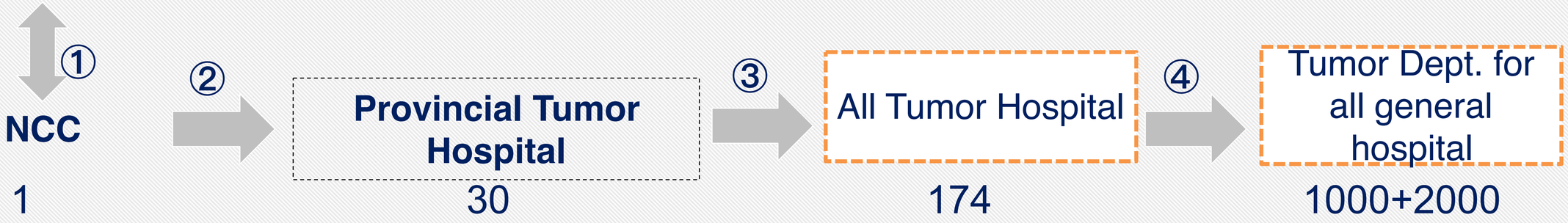
**Health Medical Big Data Oncology Committee**
**Vice President**

**Chronic Disease Prevention Alliance**

**Geriatrics Big Data Professional Committee**
**Vice President**

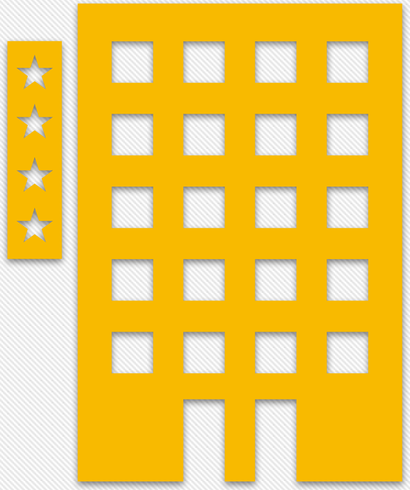
# National Cancer Center

## Data Center and Platform



# The Value of Clinical Big Data

Hospital

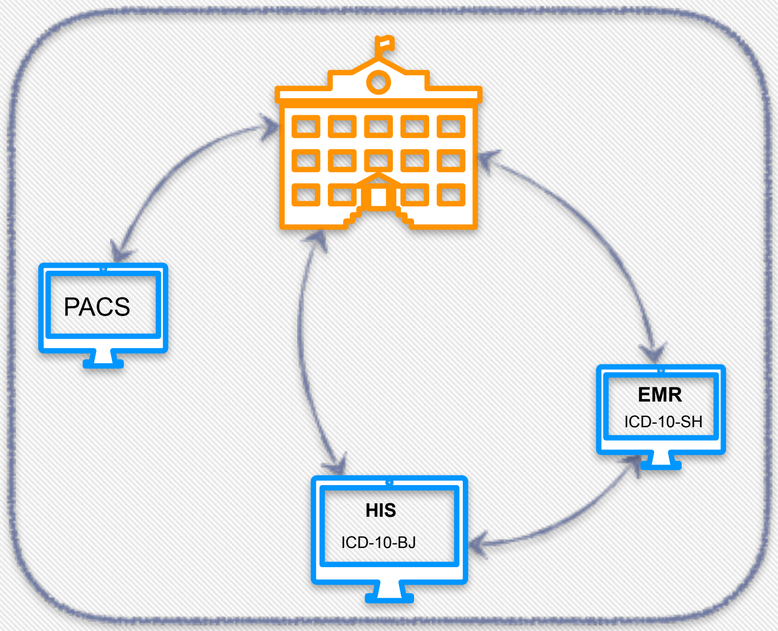


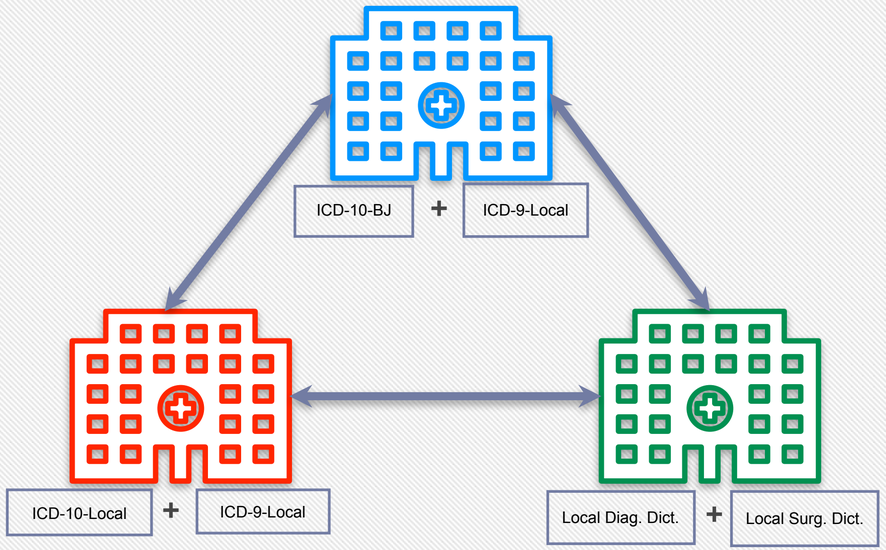
Pharma

Insurance

# Real World Data from Hospital

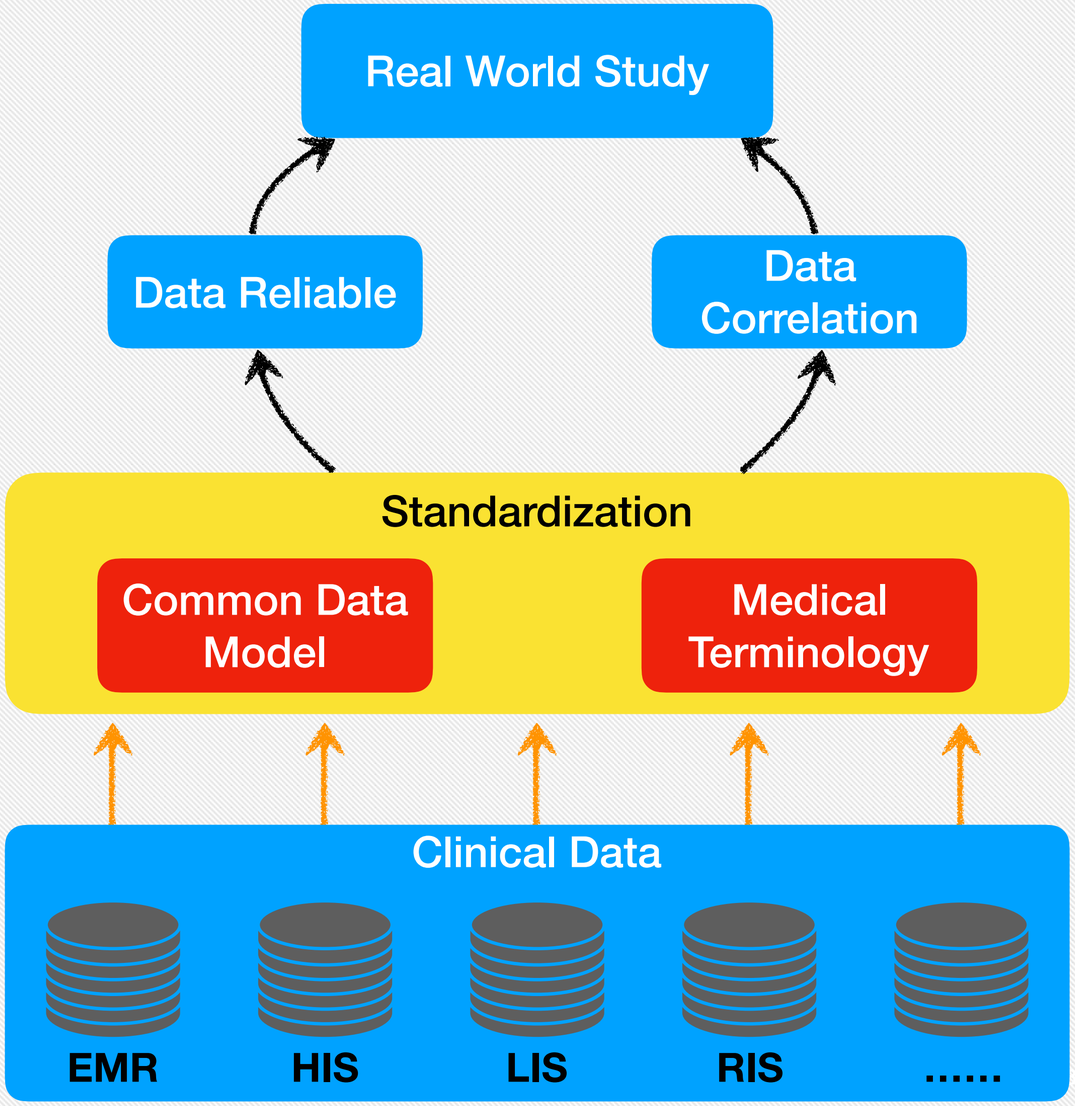
- Collection
- Unstructured Data
- Distributive distribution
- Heterogeneity



# Real World Data

- Information Model + Knowledge Model
- Data quality:
  - Data Correlation：
    - Data is collected not for research
    - Need semantic standardization
  - Data Reliable：
    - Accuracy
    - Integrity
    - Control of Bias

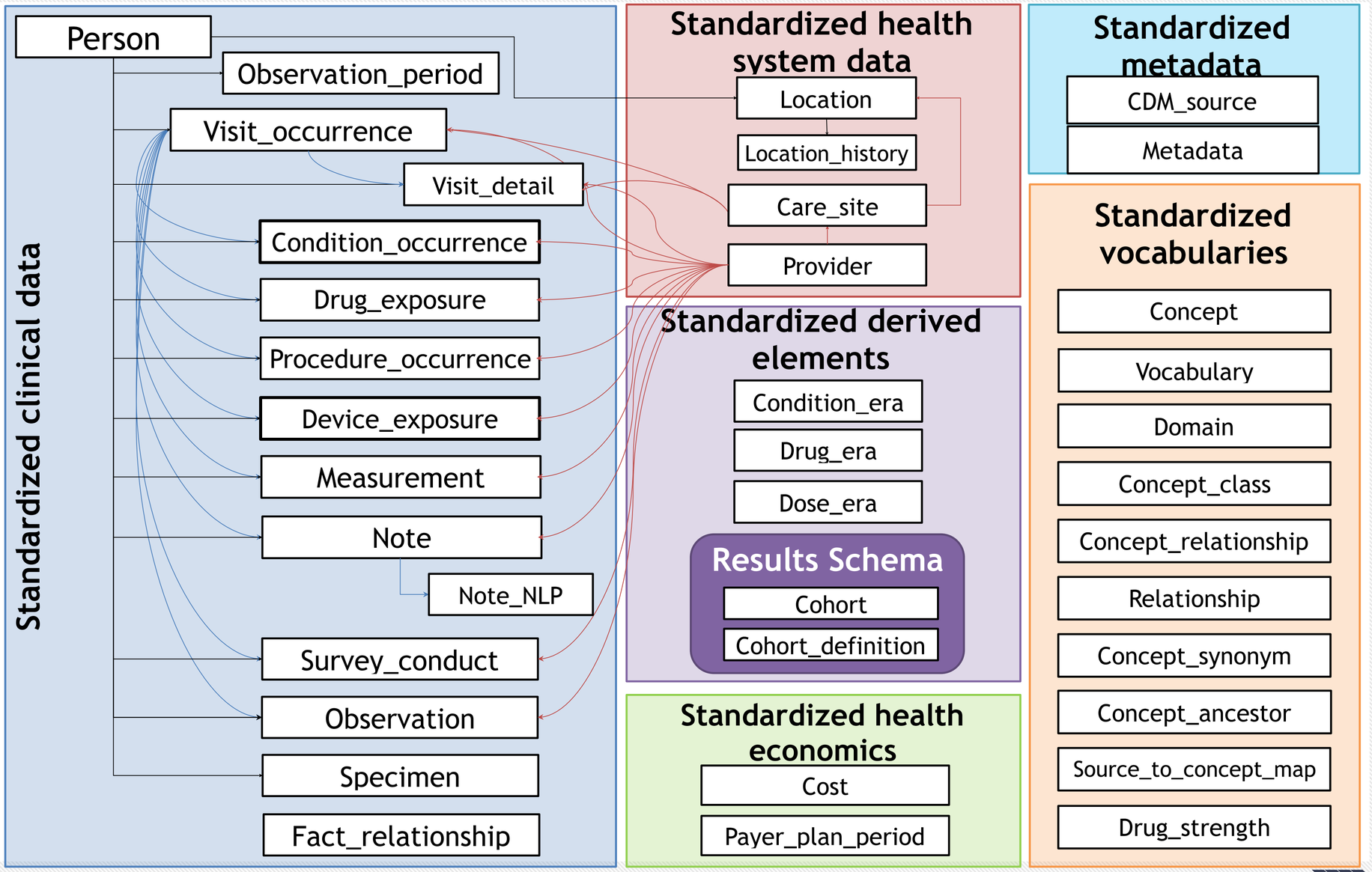



# OHDSI CDM

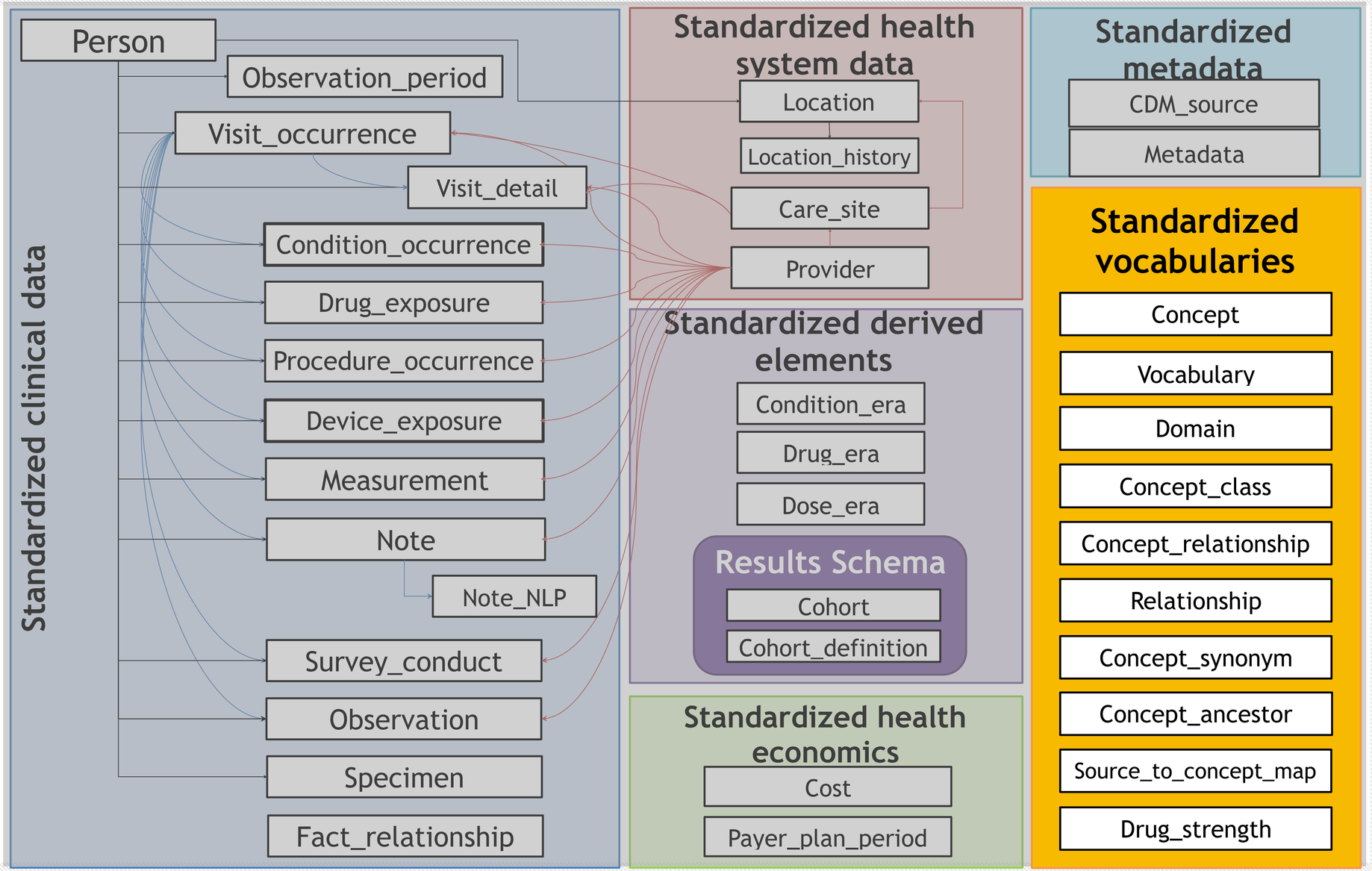
OHDSI=Observational Health Data Science and Informatics
CDM=Common Data Model

# OHDSI: Global group

# OHDSI CDM 6.0


Standardized clinical data
Person
Observation_period
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Note
Note_NLP
Survey_conduct
Observation
Specimen
Fact_relationship
Standardized health system data
Location
Location_history
Care_site
Provider
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results Schema
Cohort
Cohort_definition
Standardized health economics
Cost
Payer_plan_period
Standardized metadata
CDM_source
Metadata
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength

# OHDSI CDM 6.0


Standardized clinical data
Person
Observation_period
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Note
Note_NLP
Survey_conduct
Observation
Specimen
Fact_relationship
Standardized health system data
Location
Location_history
Care_site
Provider
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results Schema
Cohort
Cohort_definition
Standardized health economics
Cost
Payer_plan_period
Standardized metadata
CDM_source
Metadata
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength

# Background

- OHDSI Common Data Model (CDM) will be used for the big data platform (Vinci) of DCH;
- The data from the tumor hospitals would be transferred and stored in OHDSI CDM;
- LOINC is the standard vocabulary of OHDSI CDM for the domain of lab test;
-

# Terminology Mapping for LOINC

- LOINC is the standard vocabulary for lab test in OHDSI CDM.
- Prof. Zhang Lin translated the terms of LOINC to Chinese
- Six axes in LOINC: Component, Property, Timing, System, Scale and Method
- One to three axes information could be collected from information system of tumor hospitals
  - Component, System (sample)
  - single axial data comes from NLP
- It is difficult to exactly map the terms to LOINC

# Objective

- To compare the mapping for lab test name from the tumor hospitals of China to LOINC and SNOMED CT

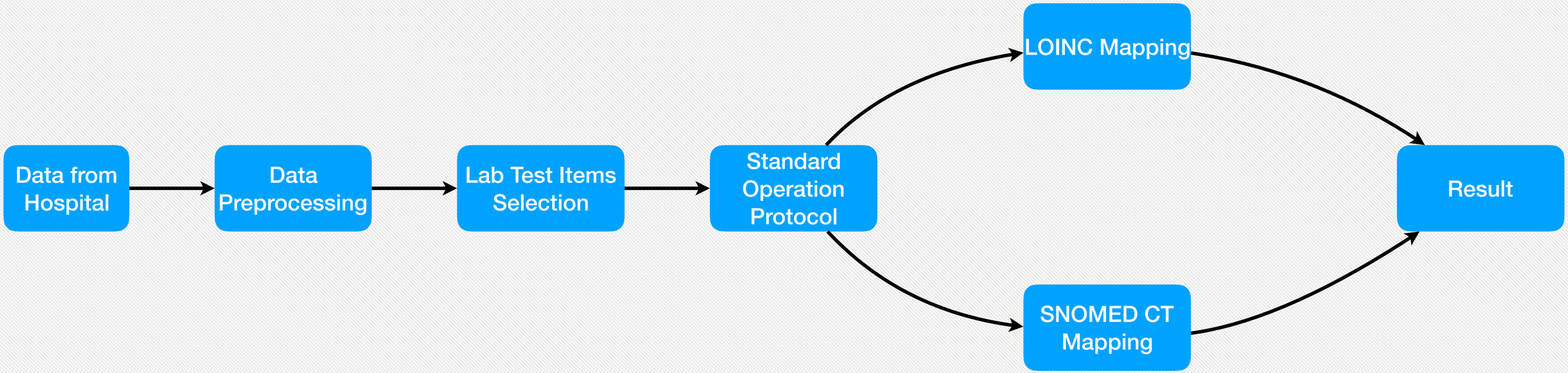
# Methods

- 4451 laboratory terms were collected from 4 tumor hospitals
- The cumulative frequency of all terms is 59285236.
- After sorting by frequency, we selected 643 high frequency items with 80% frequency to map to SNOMED
- Relma 6.22 was used for the mapping of lab test name to LOINC, language set to Chinese
- SNOMED International Browser was used for the mapping of lab test name to SNOMED CT
- Two terminologists manually mapped the 643 terms to LOINC and SNOMED CT, among which there were 100 laboratory terms were specifically related to tumor test;
- Then a team of two senior terminologists checked the mapping result.
-

# Data Description

| Data Source | No. Of Lab Test Items | Total Frequency |
|---|---|---|
| Tumor Hospital 1 | 1370 | 9127184 |
| Tumor Hospital 2 | 1588 | 22986357 |
| Tumor Hospital 3 | 843 | 7375863 |
| Tumor Hospital 4 | 1339 | 12847618 |

# Workflow for data processing



# Result

- Among the lab test terms, 308 (47%) were complete matched to LOINC and 335 concepts (53%) unmatched.
- Most terms (92%) were mapped to SNOMED CT,
  - including 180 concepts (28%) mapped to pre-coordinated expression,
  - 411 terms (64%) mapped to post-coordinated expression,
  - 52 terms (8%) unmapped.

# Discussion

- Due to the lack of information, Chinese Lab test items were hard to map to LOINC exactly;
- It is needed to find a way to map the terms to a standard vocabulary in order to perform;
- Post-Coordination might be a way to solve this issue.
- In order to solve this problem, we are trying to authoring local term with two or three attributes to represent the clinical meaning
- Carbohydrate antigen 125 (Procedure) and Cancer Antigen 125 (Substance)
  - Carbohydrate antigen 125 (Procedure) is inactive
  - Cancer Antigen 125 (Substance) is a “Substance”

# Result for SNOMED CT and LOINC Mapping

| Chinese Source Term | SPECIMEN | Source Unit | SCT ID | SCT Name | LOINCID | Component | Property | Time_ Aspect | System | Scale | Method | Class | SIRANK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 尿β2-微球蛋白 Beta 2 Microglobulin | 尿液 Urine | mg/l | 408227007 | Urine beta 2 microglobulin level (procedure) | 1953-9 | Beta-2-Microglobulin | MCnc | Pt | Urine | Qn | — | CHEM | 0 |
| | | | | | 83077-8 | Beta-2-Microglobulin | MCnc | Pt | Urine | Qn | IA | CHEM | 0 |
| 乳酸脱氢酶 Lactate Dehydrogenase | 血清 Serum | iu/l | 313854008 | Serum lactate dehydrogenase measurement (procedure) | 14804-9 | Lactate dehydrogenase | CCnc | Pt | Ser/Plas | Qn | Reaction: lactate to pyruvate | CHEM | 0 |
| | | | | | 14805-6 | Lactate dehydrogenase | CCnc | Pt | Ser/Plas | Qn | Reaction: pyruvate to lactate | CHEM | 0 |
| 甲胎蛋白[定量] Alpha Fetoprotein | 血清 Serum | ng/ml | 104404005 | Alpha-1-fetoprotein measurement, serum (procedure) | 1834-1 | Alpha-1-Fetoprotein | MCnc | Pt | Ser/Plas | Qn | — | CHEM | 386 |
| | | | | | 83073-7 | Alpha-1-Fetoprotein | MCnc | Pt | Ser/Plas | Qn | — | CHEM | 0 |
| 游离轻链k Free Light Chain K | 血清 Serum | mg/l | 444307003 | Detection of ordinal level of free immunoglobulin light chain in serum or plasma specimen (procedure) | 36916-5 | Immunoglobulin light chains.kappa.free | MCnc | Pt | Ser | Qn | — | CHEM | 594 |
| | | | | | 80515-0 | Immunoglobulin light chains.kappa.free | MCnc | Pt | Ser | Qn | Nephelometry | CHEM | 0 |

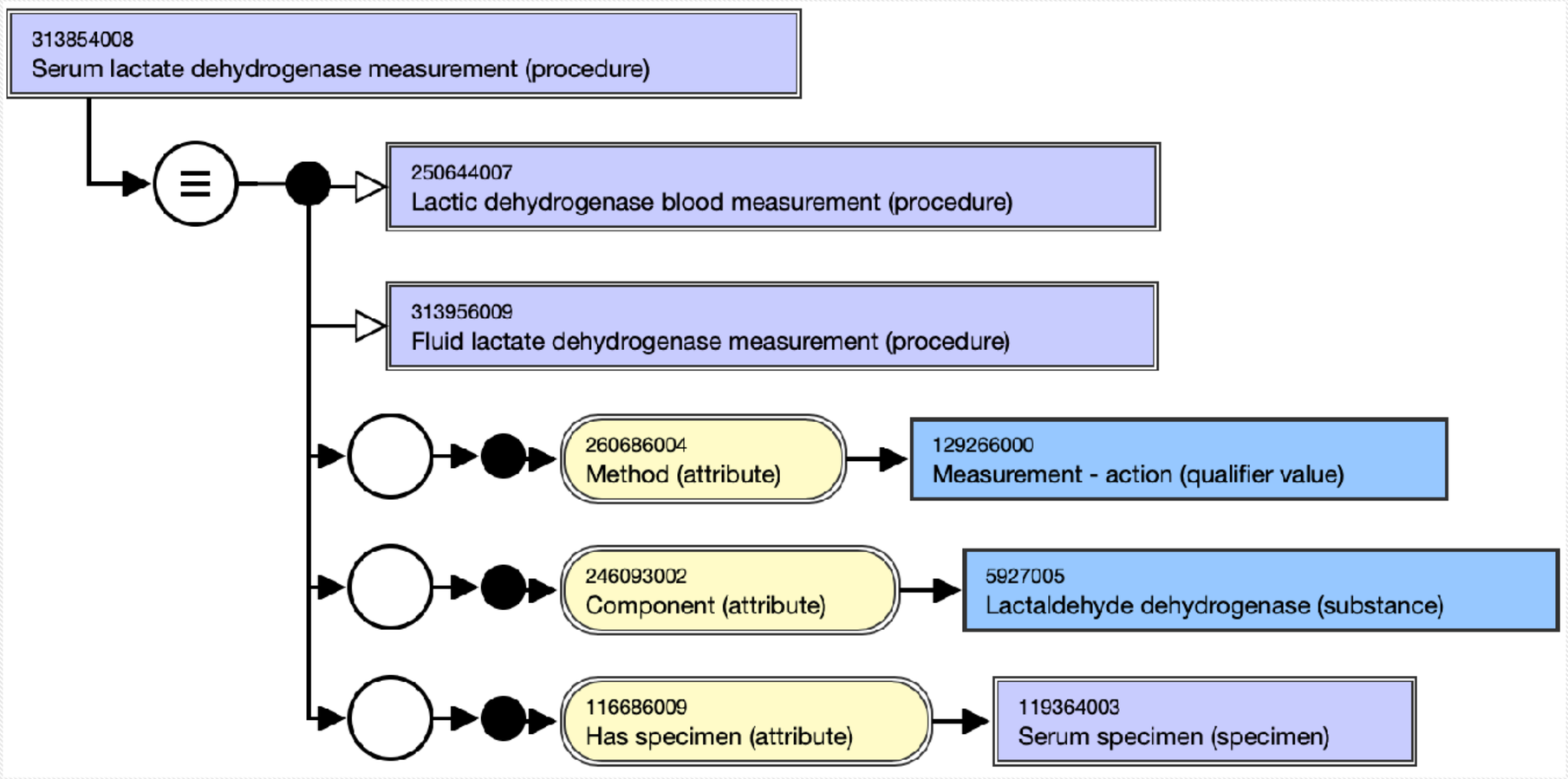
Exact Match  Special Exact Match  Supertype Match  No Match

# Result for SNOMED CT and LOINC Mapping

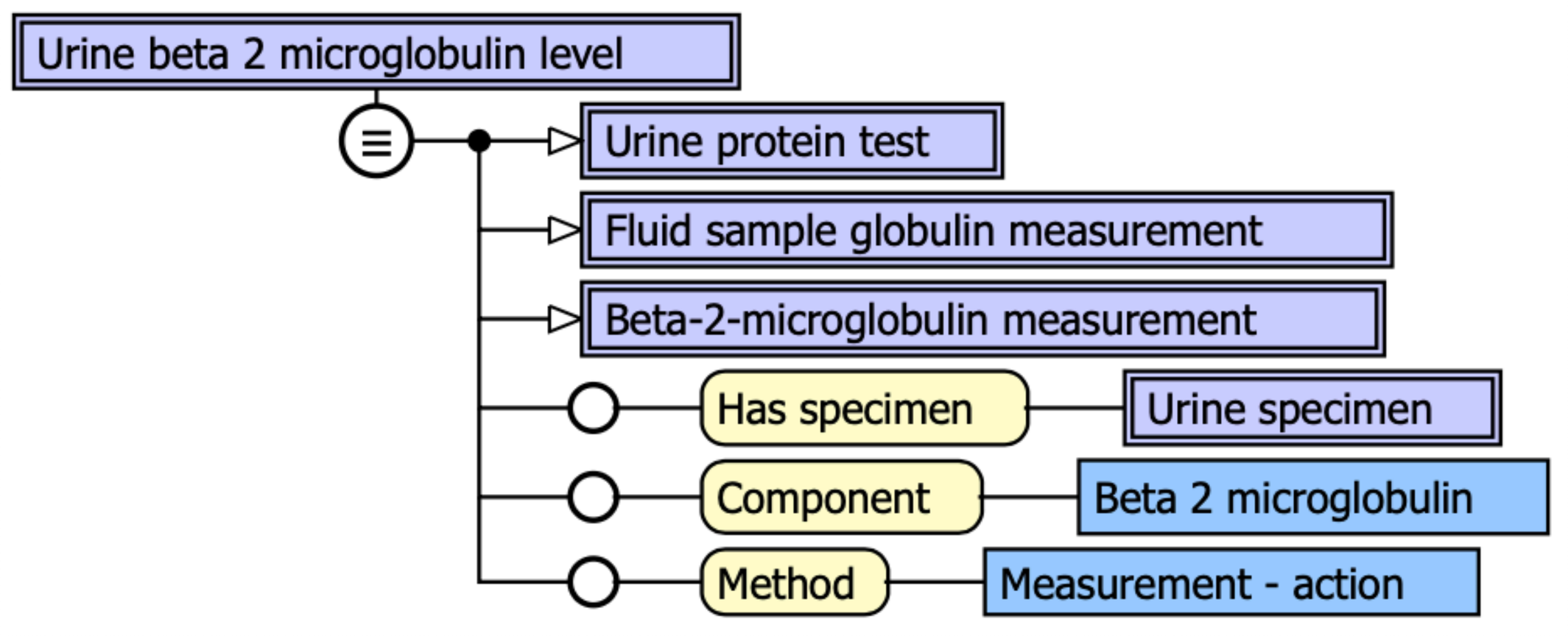
| Chinese Source Term | Source Sample term_SPECIMEN | Source Unit | SCT ID | SCT Name | LOINCID | Component | Property | Time_ Aspect | System | Scale | Method | Class | SIRANK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 癌胚抗原 Carcinoembryonic antigen | 血清 Serum | ng/ml | 60267001 | Carcinoembryonic antigen measurement (procedure) | 2039-6 | Carcinoembryonic Ag | MCnc | Pt | Ser/Plas | Qn | — | CHEM | 312 |
| | | | | | 83085-1 | Carcinoembryonic Ag | MCnc | Pt | Ser/Plas | Qn | IA | CHEM | 0 |
| 糖抗原125 Carbohydrate antigen 125 | 胸腹水 Pleural fluid/ Ascites | u/ml | 80529009 | CA 125 measurement (procedure) | 11210-2 | Cancer Ag 125 | ACnc | Pt | Body fld | Qn | — | CHEM | 0 |
| | | | | | 15156-3 | Cancer Ag 125 | ACnc | Pt | Body fld | Qn | Dilution | CHEM | 0 |
| 糖抗原199 Carbohydrate Antigen 199 | 血清 Serum | u/ml | 40939009 | Cancer antigen 19-9 measurement (procedure) | 24108-3 | Cancer Ag 19-9 | ACnc | Pt | Ser/Plas | Qn | — | CHEM | 677 |
| | | | | | 83084-4 | Cancer Ag 19-9 | ACnc | Pt | Ser/Plas | Qn | IA | CHEM | 0 |

Exact Match  Special Exact Match  Supertype Match  No Match

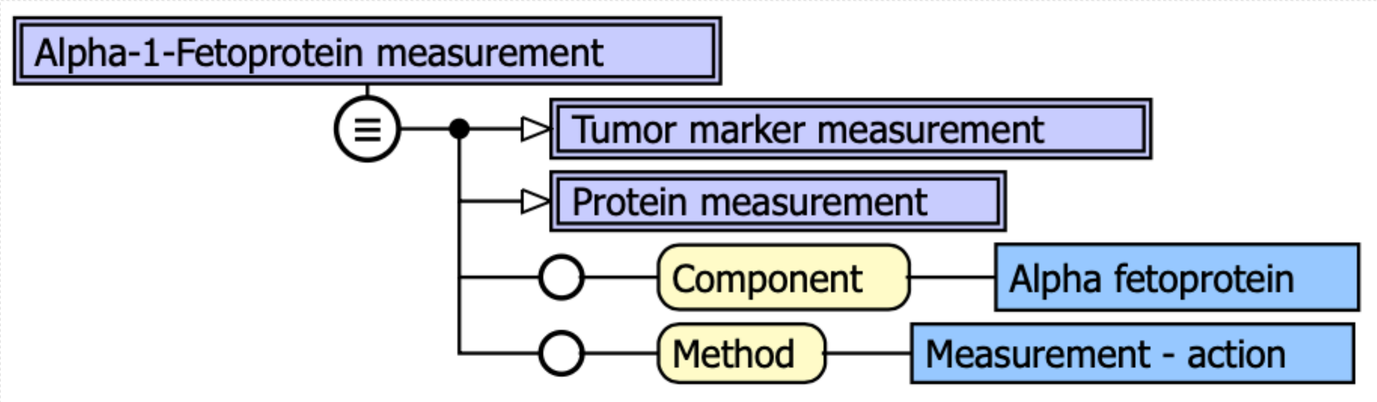
# Serum lactate dehydrogenase measurement (procedure)

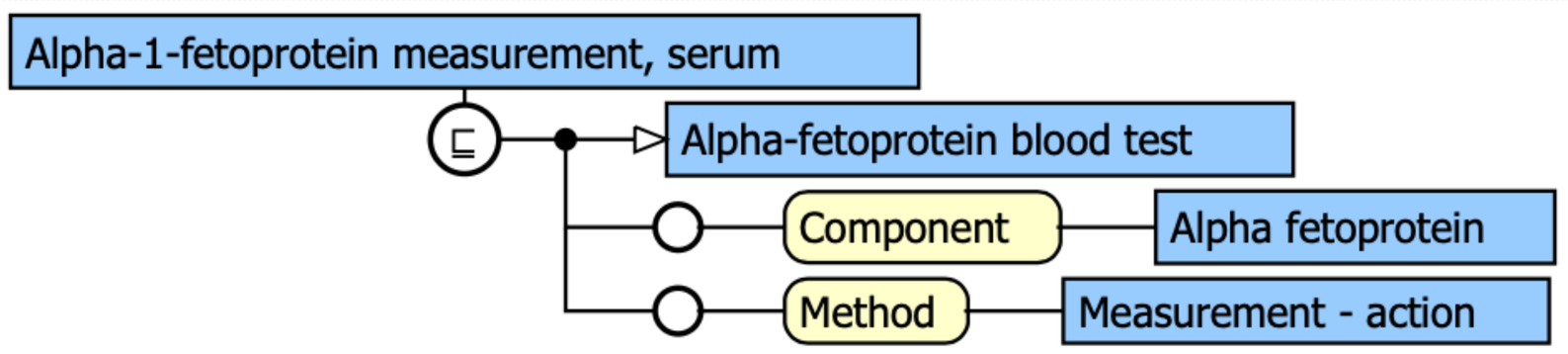


# 尿β2-微球蛋白-Urine beta-2 microglobulin



# The value of concept model for AFP measurement in serum



# Result for SNOMED CT Mapping — Post-Coordination

| Chinese Source Term | SNOMED CT Post-Coordination Expression |
| --- | --- |
| 甲胎蛋白[定量]<br>Alpha Fetoprotein | 363787002\|Observable entity (observable entity): 370130000 \| Property (attribute) \|=118539007 \| Mass concentration (property) (qualifier value) \|,704327008 \| Direct site (attribute) \|=119364003 \| Serum specimen (specimen) \|,246093002 \| Component (attribute) \|=49944008 \| Alpha fetoprotein (substance) \|,704319004 \| Inheres in (attribute) \|=67922002 \| Serum (substance) \| |
| 糖抗原125<br>CA 125 | 363787002\|Observable entity (observable entity): 704327008\| Direct site (attribute) \|=309051001 \| Body fluid sample (specimen) \|,246093002\| Component (attribute) \|=103084007 \| Cancer antigen 125 (substance) \|,370130000 \| Property (attribute) \|=118569000 \| Arbitrary concentration (property) (qualifier value) \|,\| Inheres in (attribute) \|= 32457005 \| Body fluid (substance) \| |
| 糖抗原199<br>CA 199 | 363787002\|Observable entity (observable entity): 704327008\| Direct site (attribute) \|=119364003 \| Serum specimen (specimen) \|,246093002\| Component (attribute) \|=103086009 \| Cancer antigen 19-9 (substance) \|,370130000 \| Property (attribute) \|=118569000 \| Arbitrary concentration (property) (qualifier value) \|,\| Inheres in (attribute) \|= 67922002 \| Serum (substance) \| |

# Terminologists of Digital China Health

Fengxiang Chang

Ying Zhang

Fang Wang

Chenghuan Ding

Zhi Wang

Yishang Wang

Kuangyu Ma

# 谢　谢