NKOS Consolidated Workshop 2023
Nov. 9. Daegu, South Korea

Creating a Biomedical Data Standardization Ecosystem: Ontologies, Metadata, and Applications for Effective Data Management

Xiaolin Yang

yangxl@pumc.edu.cn

Chinese Academy of Medical Sciences & Peking Union Medical College

BioMedical Branch of China National Population Health Data Center（BMICC)

2023-11-9

Outline

- **Goals and status quo of scientific data management**
- **Biomedical Data Standards Environment**
- **Standard resource construction**
- **Summary**

SCIENTIFIC DATA

OPEN

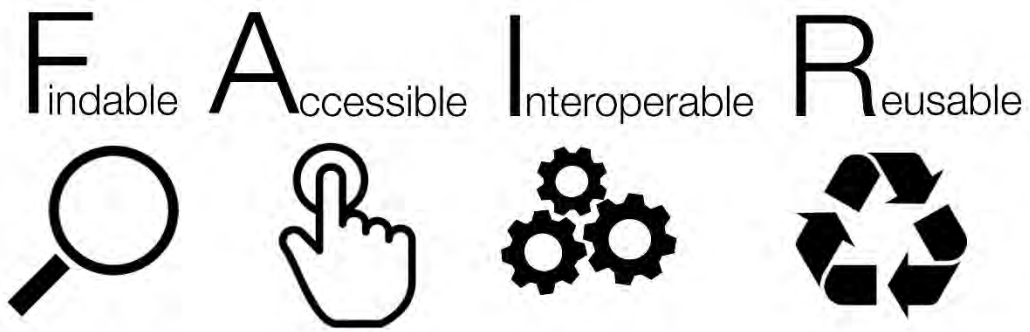
SUBJECT CATEGORIES

» Research data

» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

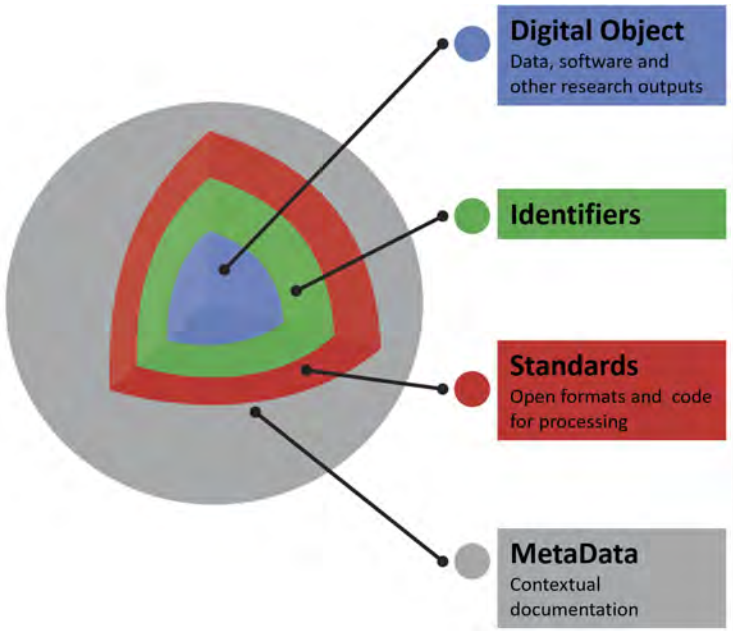
Mark D. Wilkinson *et al.*#



FAIR Principles put specific emphasis on enhancing the ability of **machines to automatically find and use the data …**

Sci Data. 2016; 3: 160018.

Details of the FAIR Principles



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Modified from 'Turning FAIR into reality'

Scientific data, 2016, 3(1): 1-9.

Scientific data Content standards

Format

Conceptual model, schema, exchange formats etc

- Define the structure and interrelation of information, and the transmission format
- e.g. FASTA, VCF

Terminology

Controlled vocabularies, taxonomies, thesauri, ontologies etc.

- Provide definitions and unambiguous identification or concepts and objects.
- e.g. Gene Ontology

Minimum information

Minimum information reporting requirements, checklists

- Report the same core, essential information
- o e.g. MIAME guidelines

Identifier

Formal systems for resources and other digital objects that allow their unique and unambiguous identification.

- DOI, ORCID iD Identifier Schema，QID（Wikidata Identifier）

Minimum Information About a Microarray Experiment - MIAME

MIAME describes the **Minimum Information About a Microarray Experiment** that is needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment. [Brazma et al., Nature Genetics]

The six most critical elements contributing towards MIAME are:

1. The raw data for each hybridisation (e.g., CEL or GPR files)
2. The final processed (normalised) data for the set of hybridisations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
3. The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
4. The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridisations are technical, which are biological replicates)
5. Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
6. The essential laboratory and data processing protocols (e.g., what normalisation method has been used to obtain the final processed data)

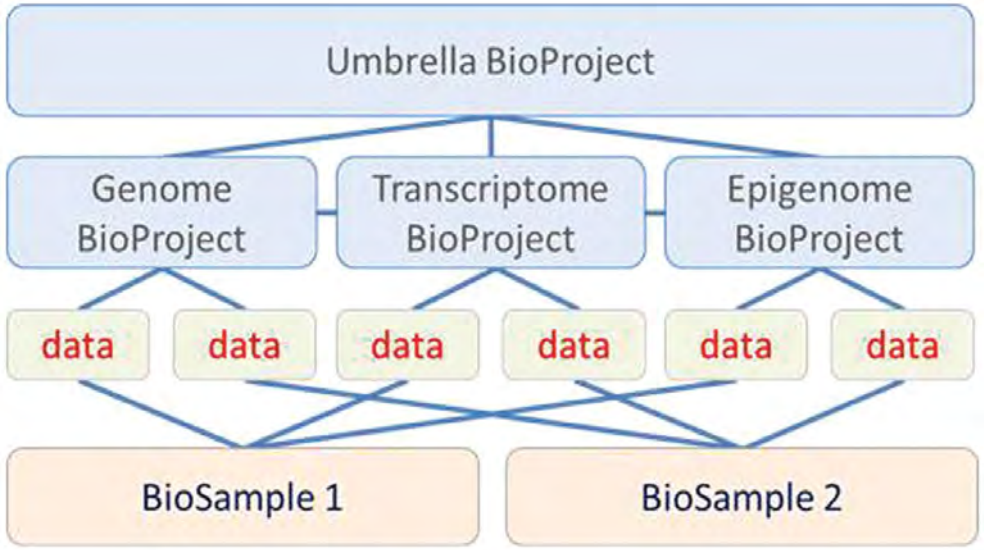
For more details, see MIAME 2.0.

MIAME does not specify a particular format, however, obviously the data are more usable, if it is encoded in a way that the essential information specified by MIAME can be accessed easily. FGED recommends the use of MAGE-TAB format, which is based on spreadsheets, or MAGE-ML.

MIAME also does not specify any particular terminology, however for automated data exchange the use of standard controlled vocabularies and ontologies are desirable. FGED recommends the use of MGED Ontology for the description of the key experimental concepts, and where possible ontologies developed by the respective community for describing terms such as anatomy, disease, chemical compounds etc (see OBO page for more detail).

Concept	CIMR[a]	MIACA	MIAME	MIAME/E	MIAME/N	MIAME/P	MIAME/To	MIAPA	MIAPE[a]	MIARE	MIFlowCy	MIGen	MIGS/MI	MIMIx	MIMPP	MINI	MIQAS	MIqPCR	MIRIAM	MISFISH	STRENDA	Row totals
Study inputs																						
Study design	●		●						●		●											17
Generic organism	●	●			●		●		●					●		●					●	18
Cells/microbes	●	●			●						●					●						7
Plant	●					●																2
Animal	●				●																	3
Mouse															●							1
Human																						3
Population																						3
Environmental sample																						2
Environment/habitat	●																					6
In silico model																						2
Study procedures																						
Organism maintenance																						1
Animal husbandry																						4
Cell/microbe culture																						4
Plant cultivation																						2
Acclimation																						1
Preconditioning/pretreatment																						7
Organism manipulation																						4
Assay inputs																						
Generic study input																						5
Organism part																						14
Organism state																						4
Organism trait																						6
Biomolecule																						3
Synthetic analyte																						4
Silencing RNA reagent																						2
Sample collection																						3
Sample processing																						12
Sample storage																						2
Sample transport																						1
Assay procedures																						
Detection/tagging/staining																						13
Generic analysis																						3
Array-based assay																						10
Capillary electrophoresis																						3
Cell phenotyping																						2
Clinical test/examination																						3
Column chromatography																						3
Electrochemical detection																						1
Electrophysiology mensuration																						1
Enzyme activity assay																						1
Flow cytometry																						2
FTIR spectroscopy																						1
Gel electrophoresis																						2

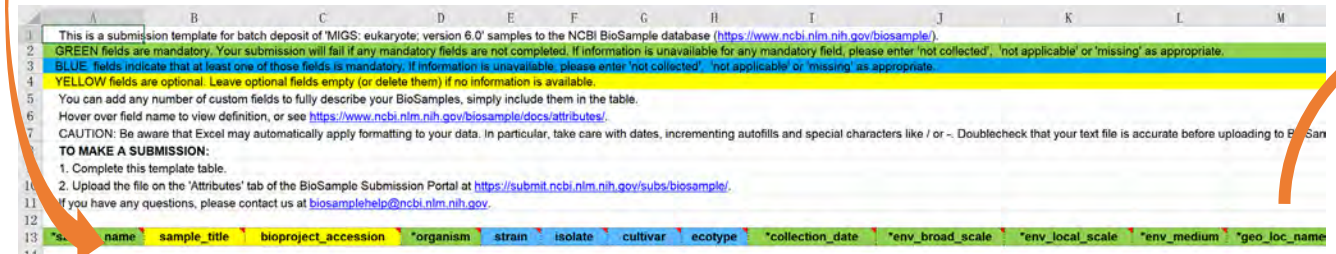
Example of NCBI primary database metadata management



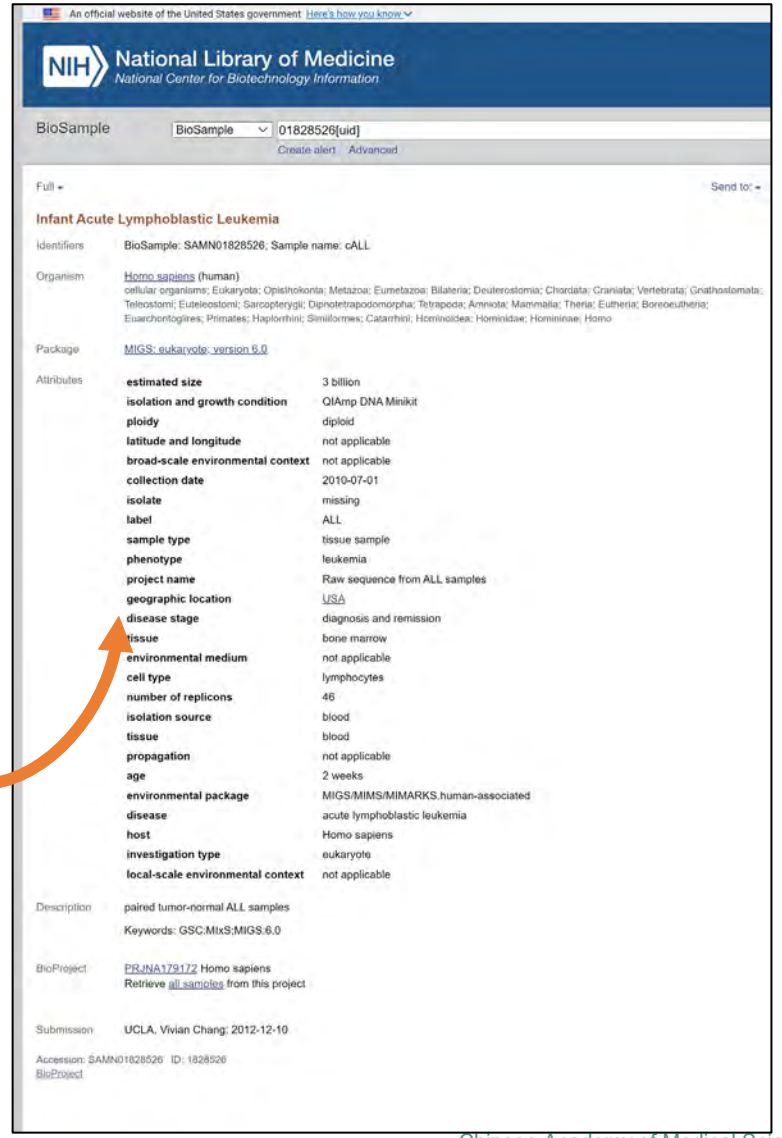
Metadata about the sample

Metadata about the study project

Schematic depicting how BioProject, BioSample and data objects can be organized and linked in NCBI *Nucleic Acids Res*, 40(D1):D57



NCBI Biosample metadata template MIGS: eukaryote, v 6.0 Package



Metadata Quality in NCBI BioSample

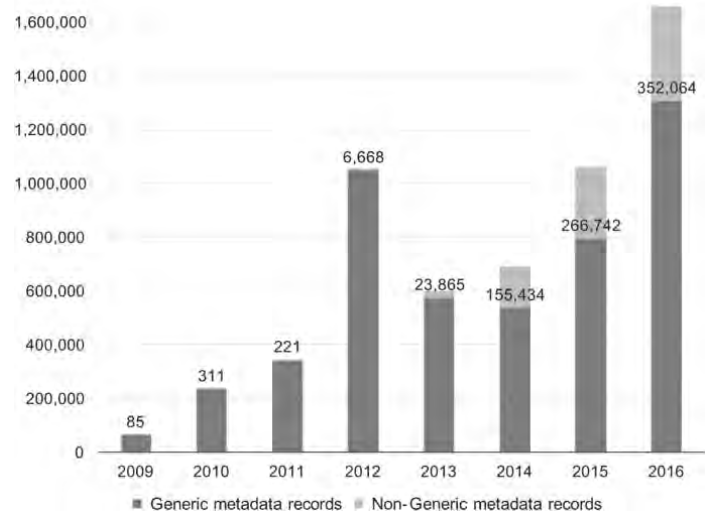


Figure 3. **Metadata submissions to NCBI BioSample from 2009–2017.**

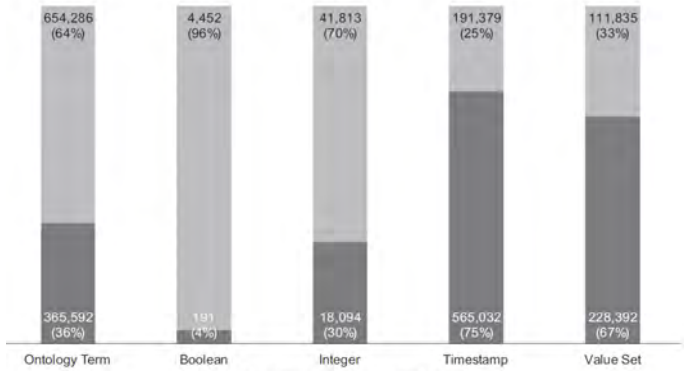


Figure 5. **Quality of attributes in packaged metadata records in NCBI BioSample.** The columns represent the metadata attribute types. Each column shows the number and percentage of metadata attributes whose values are either well-specified or invalid.

- Majority of data submitters use generic metadata templates, which lack adequate and accurate descriptions of sample attributes
  - Among the data elements (18,650) that characterize sample attributes in the BioSample database, only 2.4% (452) of the data elements (sample attributes) specified by NCBI and 97.6% are user-defined data elements
  - Only 9 of them are named using controled terminology
- Of the sample attribute values that require the use of ontology, only 36% use ontology terms in at least one value

Scientific Data 2019 Vol. 6 Issue 1 Pages 190021

Progress in Scientific Data Management and Sharing in China

- ***Measures for the Management of Scientific Data*** were issued by the Chinese government in 2018
- Scientific data generated from government-funded science and technology projects should be submitted to the designated data center.

- **National Population Health Data Center** (NPHDC, formerly named NCMI) is one of the nationally recognized data centers in biomedical and population health field
  - Main responsibility is receiving and archiving data that scientists submit
  - Developing domain data standards
  - Providing data sharing and reuse services

Population Health Data Archive (PHDA)

- https://www.ncmi.cn/

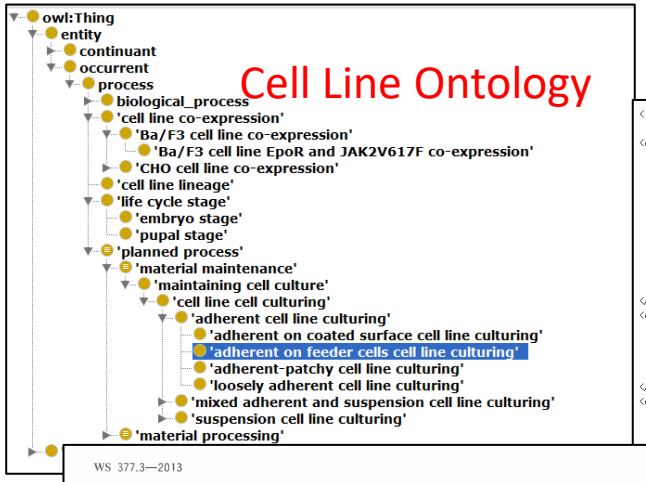
- The new system was available in 2019.
- PHDA is an infrastructure to ensure domestic researcher share their data.
- Provide a platform for querying and reusing the data

Challenges for PHDA

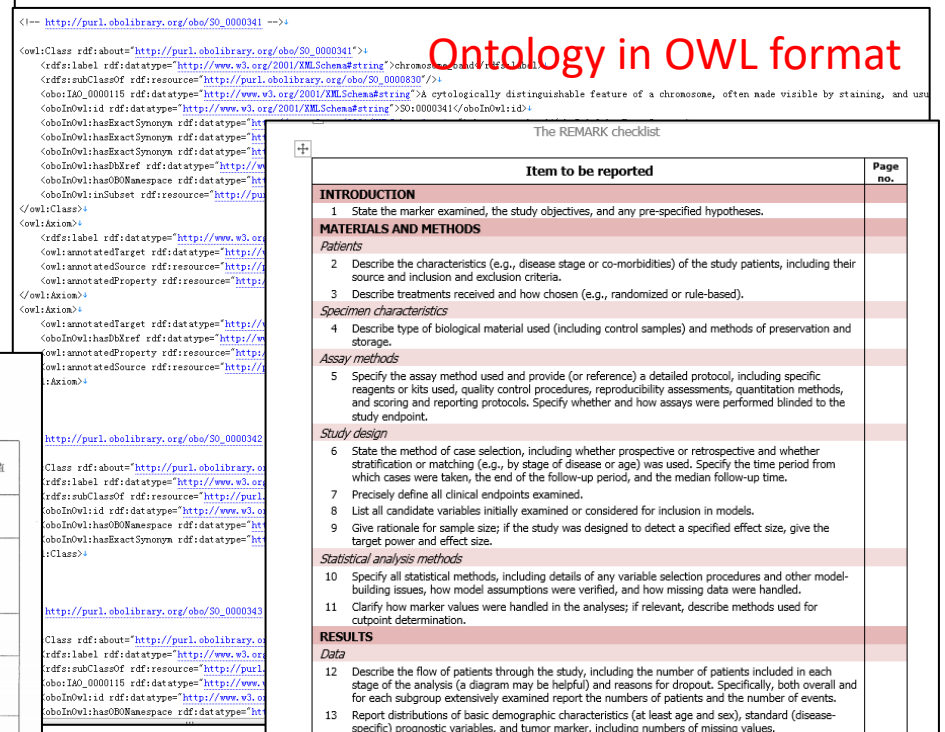
- The submitted data with complicated types from multiple sub-domains.
- Project-centric data submitting is useful for aggregating data in a short time, we still need to promote data in-depth reuse.
- Metadata submitted as data dictionary without pre-defined rules.
  - Without any semantics standard
- Difficulties in harmonizing data according to specific needs.

	字段/变量含义	是否必填	字段/变量类型	默认值	字段/变量长度	字段/变量合法取值范围	字段/变量取值范围	字段顺序
529	单核细胞百分比	否	decimal	100	8	单核细胞百分比	0~9999	528
530	嗜碱性粒细胞百分比	否	decimal	100	8	嗜碱性粒细胞百分比	0~9999	529
531	嗜酸性粒细胞百分比	否	decimal	100	8	嗜酸性粒细胞百分比	0~9999	530
532	红细胞计数	否	decimal	100	8	红细胞计数	0~9999	531
533	血红蛋白	否	decimal	100	8	血红蛋白	0~9999	532
534	红细胞平均体积	否	decimal	100	8	红细胞平均体积	0~9999	533
535	红细胞平均Hb含量	否	decimal	100	8	红细胞平均Hb含量	0~9999	534
536	红细胞平均Hb浓度	否	decimal	100	8	红细胞平均Hb浓度	0~9999	535
537	血小板计数	否	decimal	100	8	血小板计数	0~9999	536
538	血小板分布宽度	否	decimal	100	8	血小板分布宽度	0~9999	537
539	血小板比积	否	decimal	100	8	血小板比积	0~9999	538
540	血小板平均容积	否	decimal	100	8	血小板平均容积	0~9999	539
541	血小板分布宽度	否	decimal	100	8	血小板分布宽度	0~9999	540
542	红细胞比积	否	decimal	100	8	红细胞比积	0~9999	541
543	尿β2微球蛋白	否	decimal	100	8	尿β2微球蛋白	0~9999	542
544	尿肌酐	否	decimal	100	8	尿肌酐	0~9999	543
545	尿微量白蛋白	否	decimal	100	8	尿微量白蛋白	0~9999	544
546	尿微量白蛋白/尿肌酐	否	decimal	100	8	尿微量白蛋白/尿肌酐	0~9999	545
547	总蛋白	否	decimal	100	8	总蛋白	0~9999	546
548	白蛋白	否	decimal	100	8	白蛋白	0~9999	547
549	球蛋白	否	decimal	100	8	球蛋白	0~9999	548
550	白球比	否	decimal	100	8	白球比	0~9999	549
551	总胆红素	否	decimal	100	8	总胆红素	0~9999	550
552	直接胆红素	否	decimal	100	8	直接胆红素	0~9999	551
553	间接胆红素	否	decimal	100	8	间接胆红素	0~9999	552
554	谷草转氨酶	否	decimal	100	8	谷草转氨酶	0~9999	553
555	谷丙转氨酶	否	decimal	100	8	谷丙转氨酶	0~9999	554
556	谷丙转氨酶/谷草转氨酶	否	decimal	100	8	谷丙转氨酶/谷草转氨酶	0~9999	555

Data standards are difficult for domain scientists to use



Cell Line Ontology



Ontology in OWL format

WS 377.3—2013

表 3（续）

内部标识符	数据元标识符（DE）	数据元名称	定义	数据元值的数据类型	表示格式	数据元允许值
HDSB02.03.126	DE05.10.062.00	手术并发症详细描述	对患者手术并发症的详细描述	S1	AN..100	
HDSB02.03.127	DE06.00.018.00	处理及指导意见	对受检者进行计划生育技术服务后的处理及指导意见内容的详细描述	S1	AN..200	
HDSB02.03.128	DE08.10.013.00	手术机构名称	手术机构的组织机构名称	S1	AN..70	
HDSB02.03.129	DE02.01.039.00	手术医师姓名	手术医师在公安管理部门正式登记注册的姓氏和名称	S1	A..50	
HDSB02.03.130	DE06.00.109.00	随诊日期	对本人进行医学随诊当日的公元纪年日期	D	D8	
HDSB02.03.131	DE04.01.119.00	主诉	患者向医师描述的对自身本次疾病相关的感受的主要记录	S1	AN..100	
HDSB02.03.132	DE05.10.069.00	随诊检查结果	随诊时检查结果的详细描述	S1	AN..100	
HDSB02.03.133	DE06.00.007.00	带教指导	对接受相关服务的对象进行带教方面指导的具体内容	S1	AN..100	
HDSB02.03.134	DE08.10.013.00	随诊机构名称	随诊机构的组织机构名称	S1	AN..70	

Chinese Data Element Standards, PDF Format

The REMARK checklist

	Item to be reported	Page no.
INTRODUCTION		
1	State the marker examined, the study objectives, and any pre-specified hypotheses.	
MATERIALS AND METHODS		
Patients		
2	Describe the characteristics (e.g., disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.	
3	Describe treatments received and how chosen (e.g., randomized or rule-based).	
Specimen characteristics		
4	Describe type of biological material used (including control samples) and methods of preservation and storage.	
Assay methods		
5	Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.	
Study design		
6	State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.	
7	Precisely define all clinical endpoints examined.	
8	List all candidate variables initially examined or considered for inclusion in models.	
9	Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.	
Statistical analysis methods		
10	Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.	
11	Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.	
RESULTS		
Data		
12	Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.	
13	Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.	
Analysis and presentation		
14	Show the relation of the marker to standard prognostic variables.	
15	Present univariable analyses showing the relation between the marker and outcome, with the estimated effect (e.g., hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.	
16		
17		
18		
DI		
19		
20		

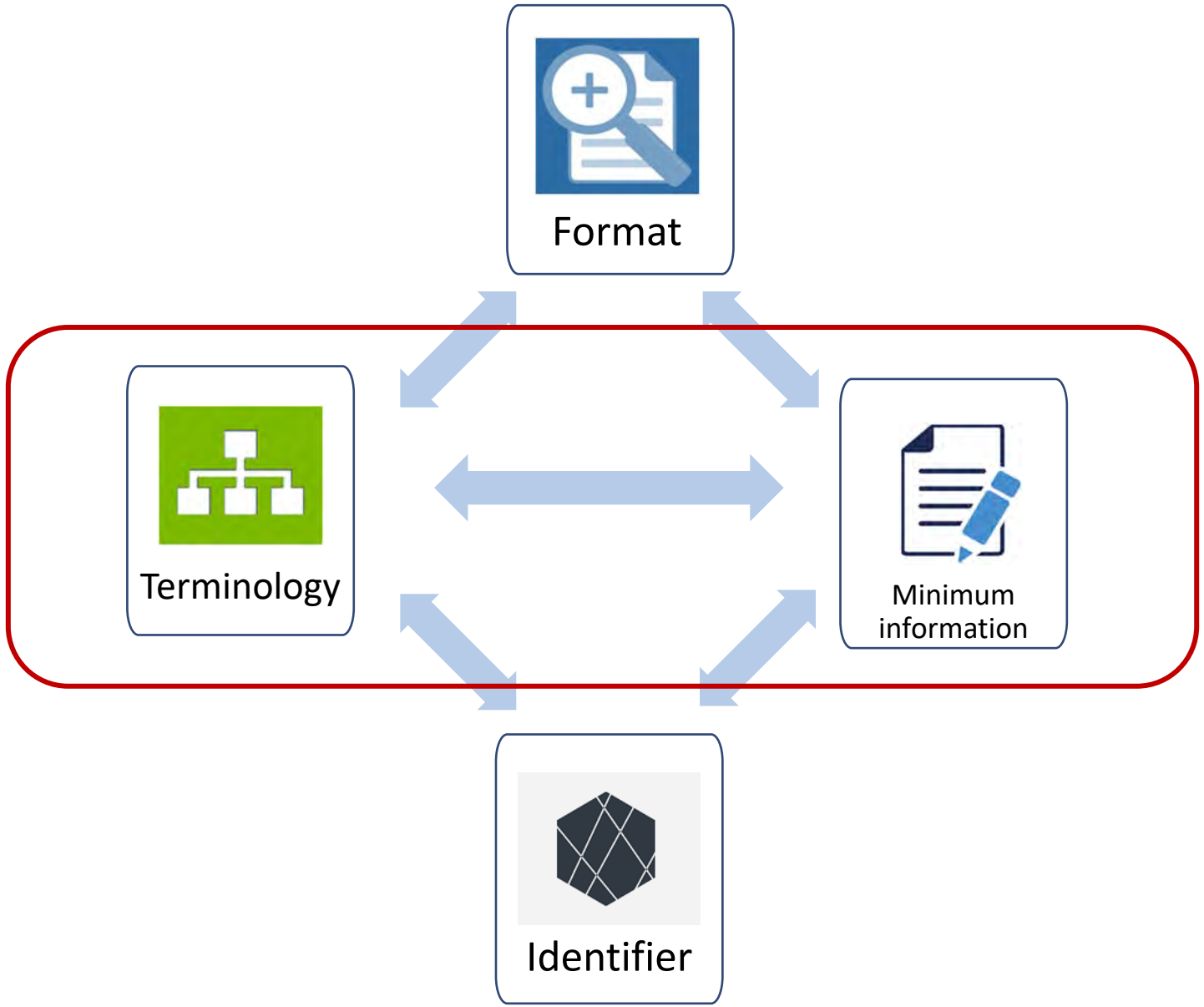
marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005; 97: 1180-1184.

Minimum information reporting requirements, checklists in WORD format

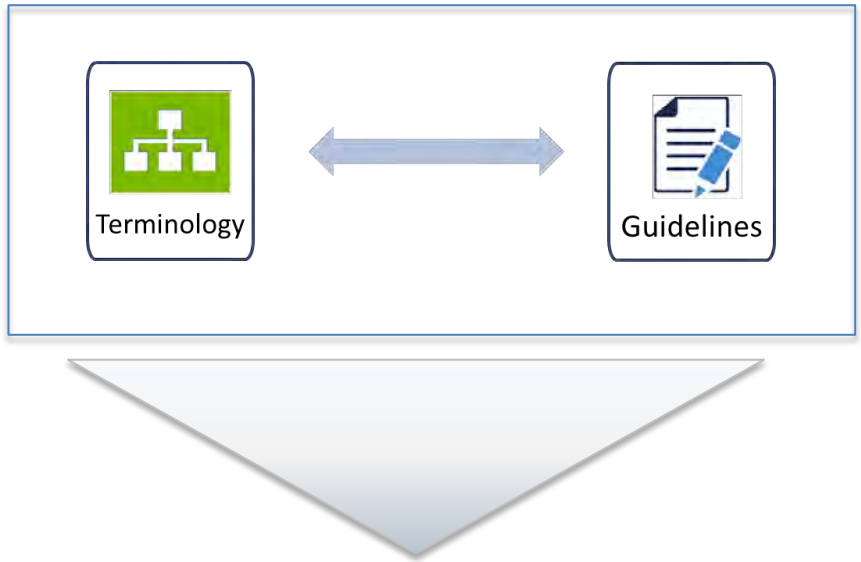
- Hard to find
- Hard to obtain
- Hard to understand
- Hard to use

Outline

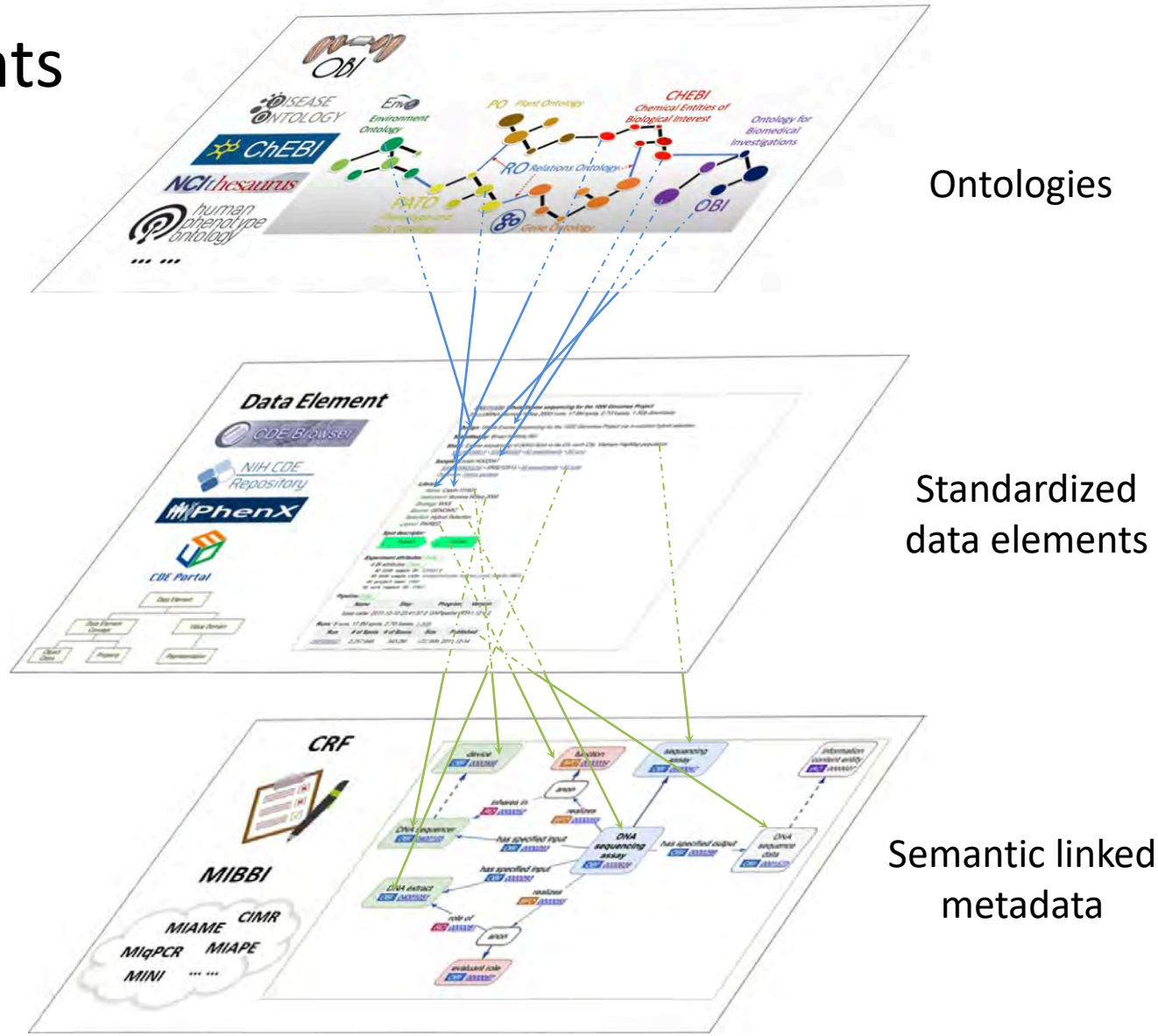
- Goals and status quo of scientific data management
- **Biomedical Data Standards Environment**
- Standard resource construction
- Summary


Format
Terminology
Minimum information
Identifier

From ontology and data elements to metadata standardization



Machine-understandable
content metadata standards

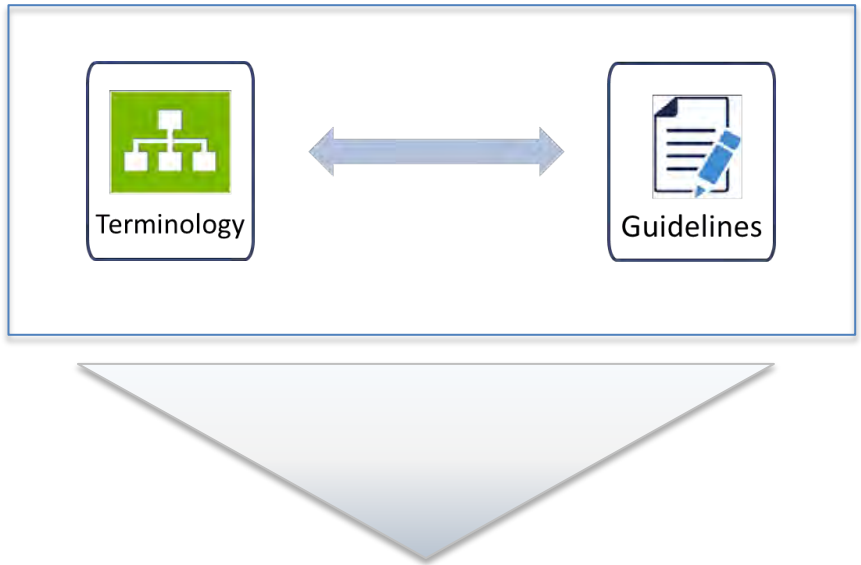


Ontologies

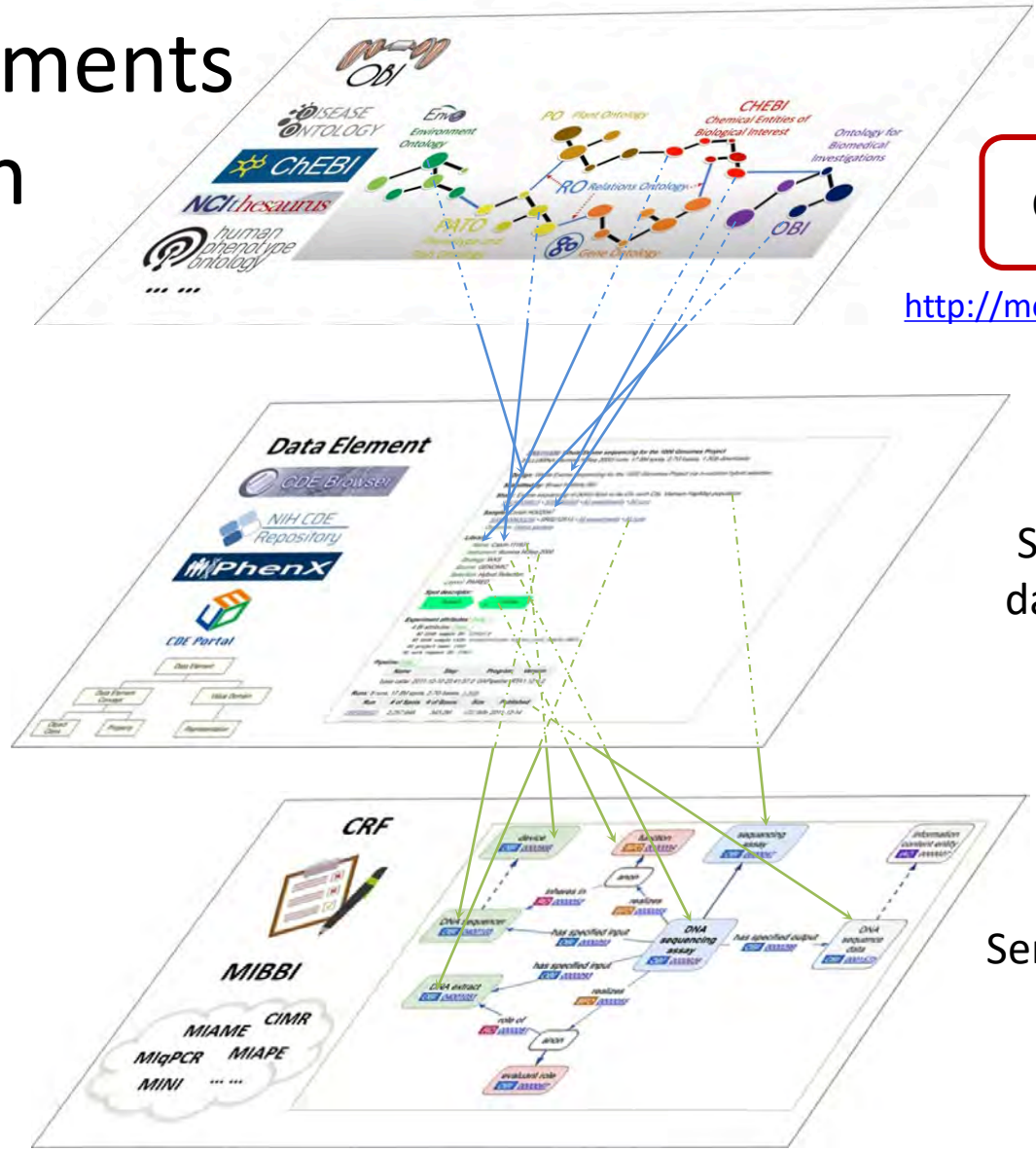
Standardized data elements

Semantic linked metadata

From ontology and data elements to metadata standardization



Machine-understandable
content metadata standards



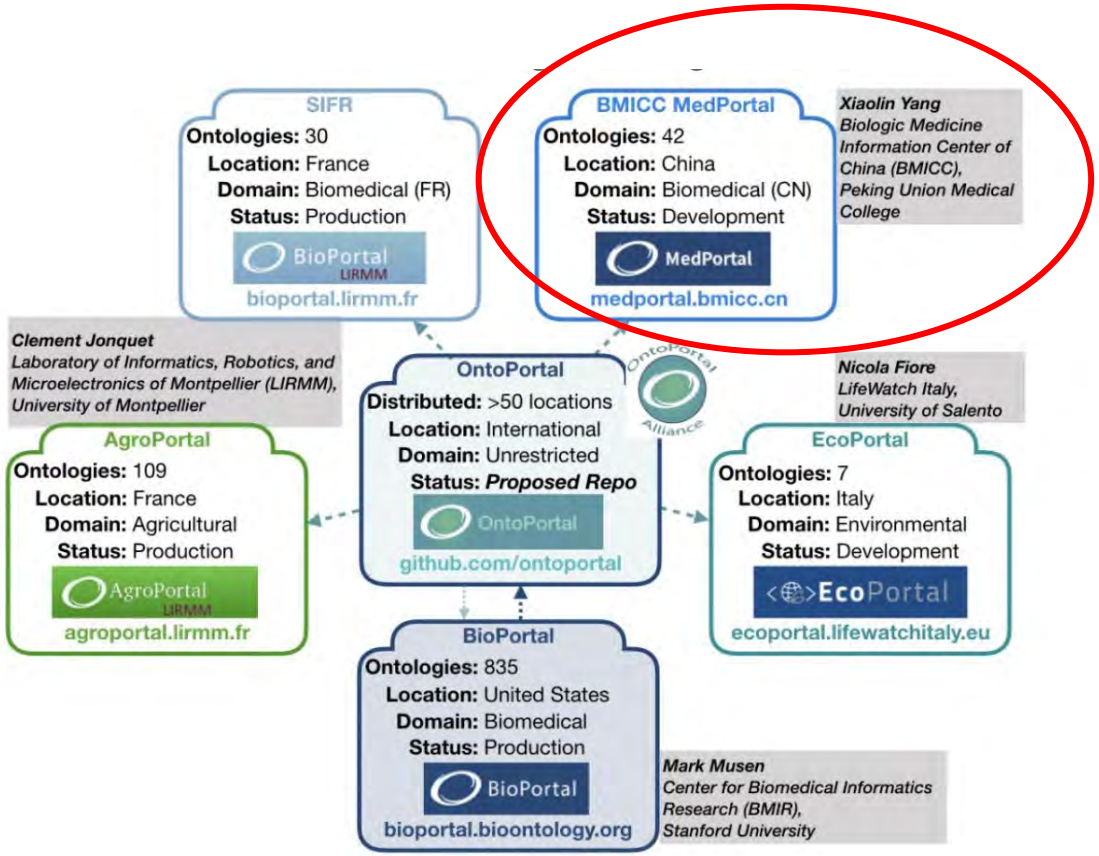
Ontologies

http://medportal.bmicc.cn

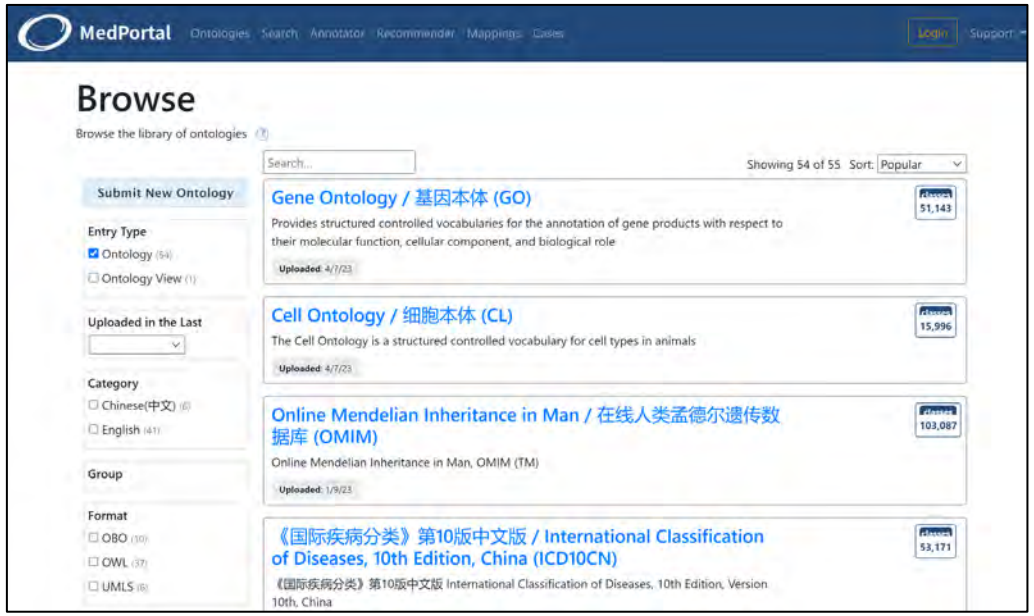
Standardized data elements

Semantic linked metadata

MedPortal – a repository of biomedical ontology resources



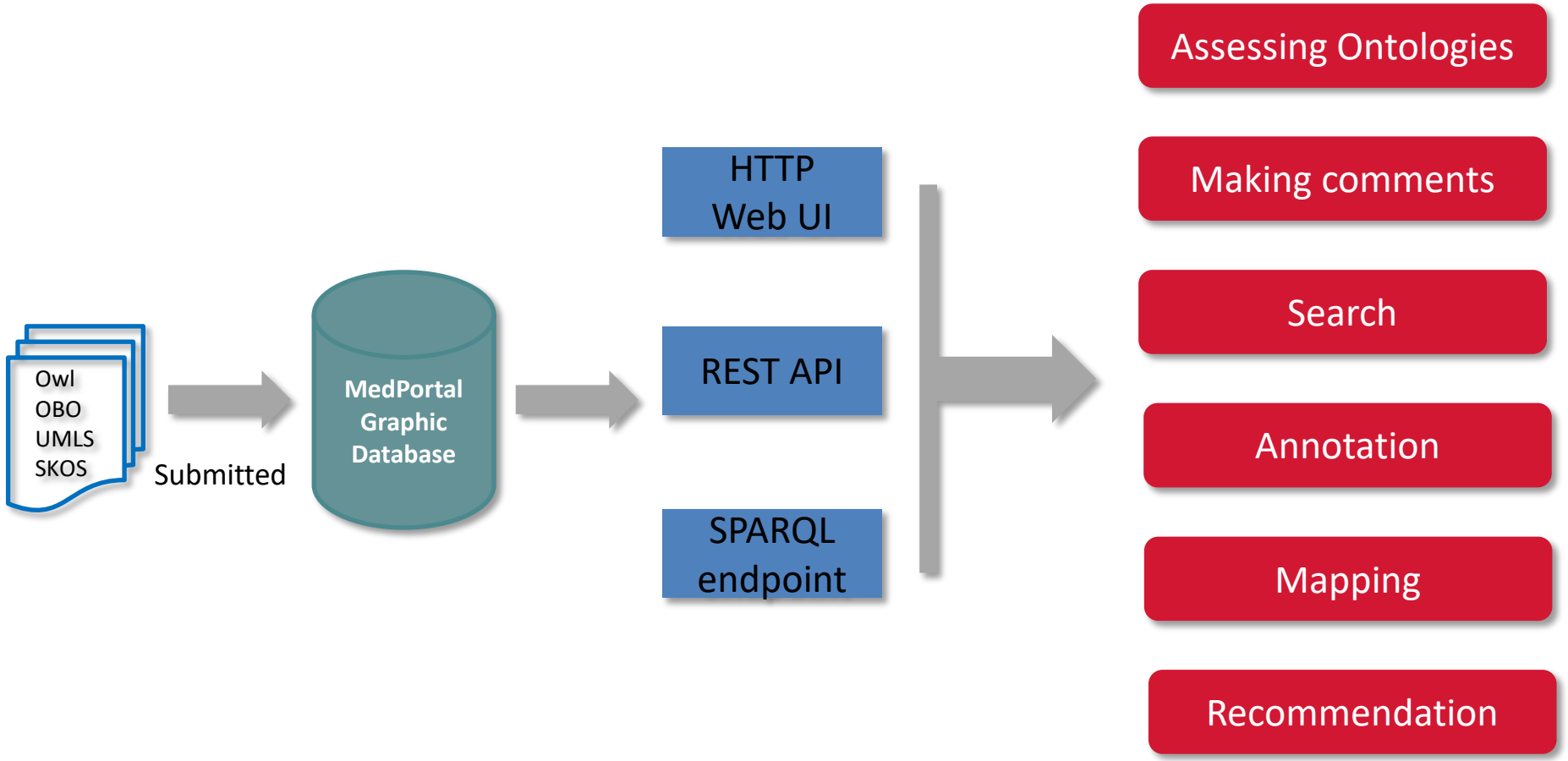
- Reuse the Ontoportal Alliance (BioPortal) framework
- Provide bilingual ontology services in Chinese and English
- Ontology services that support semantic standards



https://ontoportal.org/

URL: http://medportal.bmicc.cn/

Basic Functions of MedPortal



Semantic standardization with MedPortal

Annotator

Get annotations for biomedical text with classes from the ontologies

Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye.

Annotations

*total results **5** (direct **5** / ancestor **0** / mapping **0**)*

CLASS filter	ONTOLOGY filter	TYPE filter	CONTEXT	MATCHED CLASS	MATCHED ONTOLOGY
Melanoma	Human Phenotype Ontology	direct	**Melanoma** is a malignant ...	Melanoma	Human Phenotype Ontology
Neoplasm	Human Phenotype Ontology	direct	... a malignant **tumor** of melanocytes which ...	Neoplasm	Human Phenotype Ontology
Skin	Foundational Model of Anatomy	direct	... predominantly in **skin** but also in ...	Skin	Foundational Model of Anatomy
Intestine	Foundational Model of Anatomy	direct	... in the **bowel** and the eye.	Intestine	Foundational Model of Anatomy
Eye	Foundational Model of Anatomy	direct	... and the **eye**.	Eye	Foundational Model of Anatomy

Ontology terminology annotation

← Previous 1 2 3 4 5 6 7 8 9 ... 3132 Next →

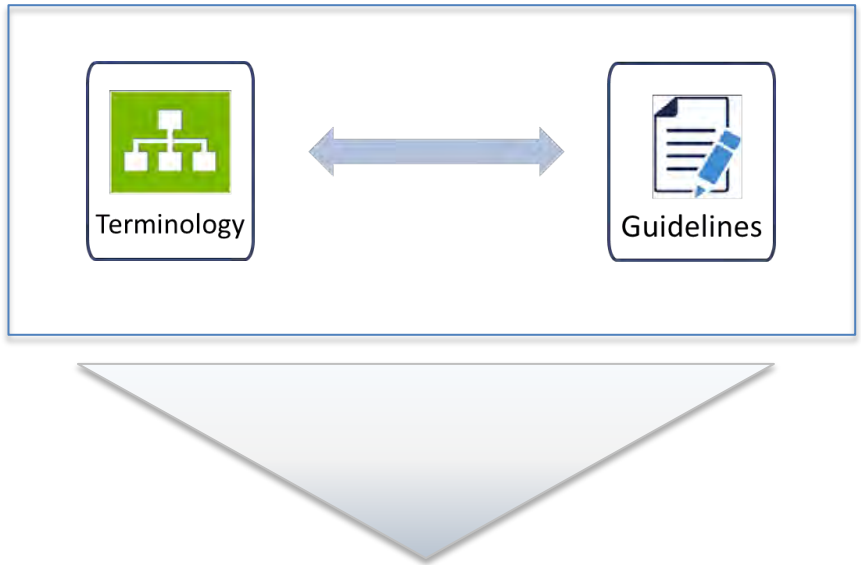
HUMAN PHENOTYPE ONTOLOGY	EXPERIMENTAL FACTOR ONTOLOGY	SOURCE
corneal neovascularization	Corneal neovascularization	LOOM
Prelingual sensorineural hearing impairment	Prelingual sensorineural hearing impairment	SAME_URI
Hyperkalemia	Hyperkalemia	SAME_URI
adrenocortical adenoma	Adrenocortical adenoma	LOOM
Decreased liver function	Decreased liver function	SAME_URI
Breech presentation	Breech presentation	SAME_URI
anxiety	Anxiety	LOOM
Oral ulcer	Oral ulcer	SAME_URI
Diffuse mesangial sclerosis	Diffuse mesangial sclerosis	SAME_URI
Talipes equinovarus	Talipes equinovarus	SAME_URI
Cyclic neutropenia	Cyclic neutropenia	LOOM
papillary thyroid carcinoma	Papillary thyroid carcinoma	LOOM
Breast mass	Breast mass	SAME_URI
superficial spreading melanoma	Superficial spreading melanoma	LOOM

Ontology terminology mapping

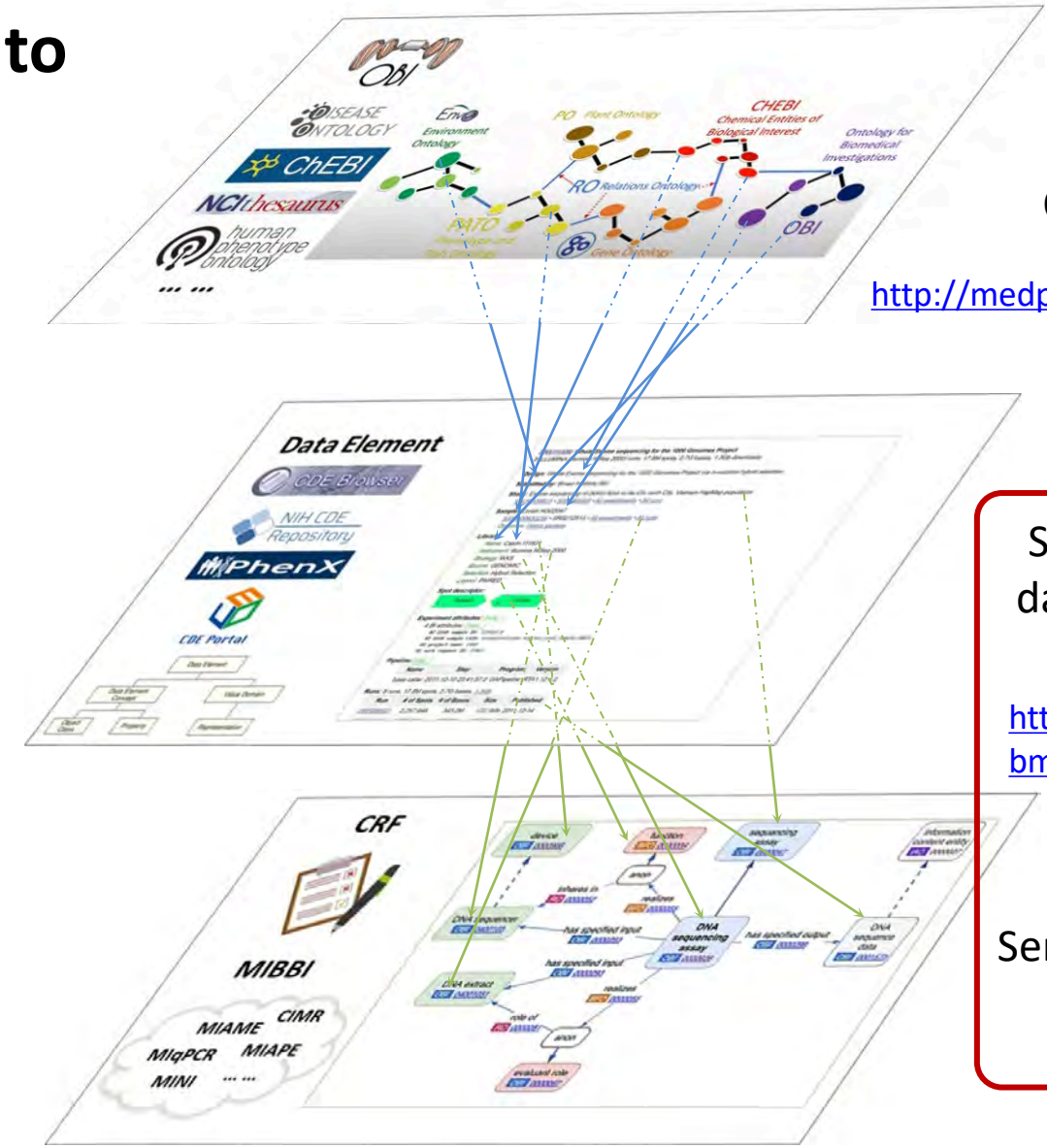
REST API

- Annotate text
  - GET POST `/annotator?text={input text}`
  - example: /annotator?text=Melanoma+is+a+malignant+tumor+of+melanocytes+which+are+found+predominantly+in+skin+but+also+in+the+bowel+and+the+eye.

From ontology and data elements to metadata standardization



Machine-understandable
content metadata standards



Ontologies

http://medportal.bmicc.cn

Standardized data elements

http://cdeportal.bmicc.cn

Semantic linked metadata

CDE Portal

- https://cdeportal.bmicc.cn/

- Provide tools for metadata compilation, registration, review, and publication for the NCMI
  - Tools for data element building
  - Tools for reusing data elements

- Provide high-quality, common data elements（Common Data Element, CDE）
  - Data elements for important research projects at home and abroad
  - Important, standardized domain data elements

Usage of CDE Portal

CDE Portal

Guide new DE generation

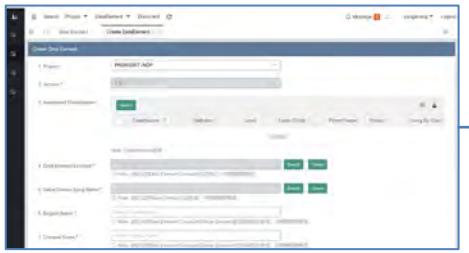
Reuse **Data Element**

Data Element Concept *from* **DE1**

Value Domain *from* **DE2**

Question *from* **DE1**

ISO/IEC 11179



Project name：XXX

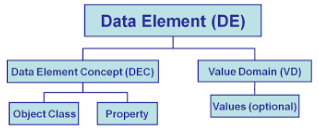
DE*1*:

DE*2*:

DE*3*:

DE*n*:

DE Database



- **CDEs and its components**
- **CDE context (project)**
- **Forms**
- **Controlled vocabulary used in CDEs**

CSV, XML *and* JSON

Download
API

Data Standardization

Support for de novo construction of DE

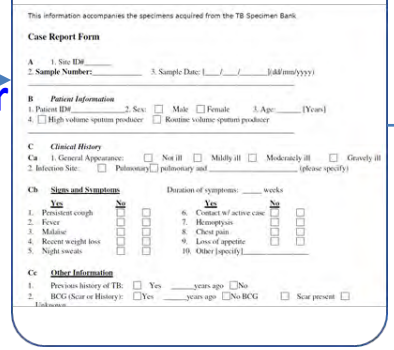
OntoPortal Alliance

MedPortal

Use ontology to descript concept of DEs and its value domains

Construct forms for data collection

Case Report Form



Data Collection

Represent Data Elements in a FAIR way

Cancer Generic Primary Tumor TNM Finding/癌症一般性原发肿瘤TNM分期 Standard

A cancer finding in the TNM system that is relevant to the diagnosis of cancer.

Preferred Question: Cancer Generic Primary Tumor TNM Finding

Value Domain: Generic Primary Tumor TNM Finding

Data Element Concept: Cancer Generic Primary Tumor TNM Finding

Public ID: DE0071220 | Version: 1.0 | Project: BMICC | workflow: Draft new | Create Date: 2021-10-16

permissible Value	Value Meaning
Any T	Any T
T0	T0 Stage Finding
T1	T1 Stage Finding
T1a	T1a Stage Finding

Implement semantic support for the definition and value of data elements

Object Class Concepts

English Name	Chinese Name	Definition	Concept Code	ontology	Primary
Cancer TNM Finding	None	A cancer finding in the TNM system that is relevant to the diagnosis of cancer.	http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C48232	NCI Thesaurus	是

Property Concepts

English Name	Chinese Name	Definition	Concept Code	Ontology	Primary
Generic Primary Tumor TNM Finding	None		http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C48885	NCI Thesaurus	是

Representation Concepts

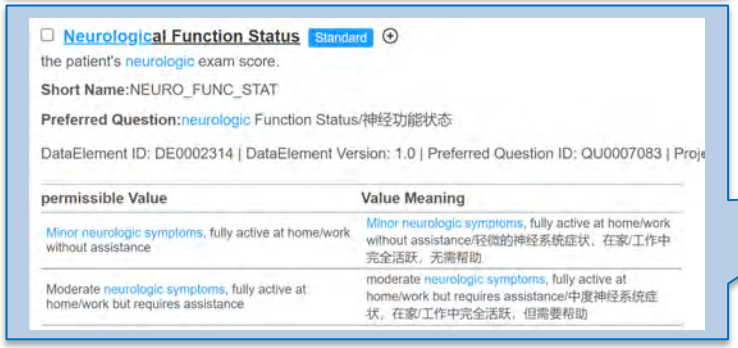
English Name	Chinese Name	Definition	Concept Code	Ontology	Primary
Generic Primary Tumor TNM Finding	None		http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C48885	NCI Thesaurus	是

Permissible Values

Public ID	Value	PV Meaning	Description	Begin Date - End Date	Project
PV0102942	Any T	Any T	No specification on the details of the primary tumor growth.	2021-10-16 -	BMICC
PV0035492	T0	T0 Stage Finding	A primary tumor TNM finding indicating that there is no evidence of primary tumor.	2021-10-16 -	NCI caDSR
PV0035493	T1	T1 Stage Finding	A clinical and/or pathologic primary tumor TNM finding indicating that the cancer is limited to the site of growth.	2021-10-16 -	NCI caDSR
PV0035494	T1a	T1a Stage Finding	A general term that refers to a TNM finding of a primary tumor limited to the site of growth. The definition of T1a TNM finding depends on the specific type of cancer that it refers to; for example, for breast cancer it refers to a primary tumor that is more than 0.1cm, but not more than 0.5 cm in greatest dimension; for kidney cancer it refers to a primary tumor that is 4 cm or less in greatest dimension; and for thyroid cancer it refers to a primary tumor that is 1 cm or less in greatest dimension.	2021-10-16 -	NCI caDSR
PV0035495	T1b	T1b Stage Finding	A general term that refers to a TNM finding of a primary tumor limited to the site of growth. The definition of T1b TNM finding depends on the specific type of cancer that it refers to; for example, for breast cancer it refers to a primary tumor that is more than 0.5 cm, but not more than 1.0 cm in greatest dimension; for kidney cancer it refers to a primary tumor that is more than 4 cm, but not more than 7 cm in greatest dimension; and for thyroid cancer it refers to a primary tumor	2021-10-16 -	NCI caDSR

urus.owl#C48885

Case investigation forms (CRFs) facilitate the reuse of data elements and improve data interoperability

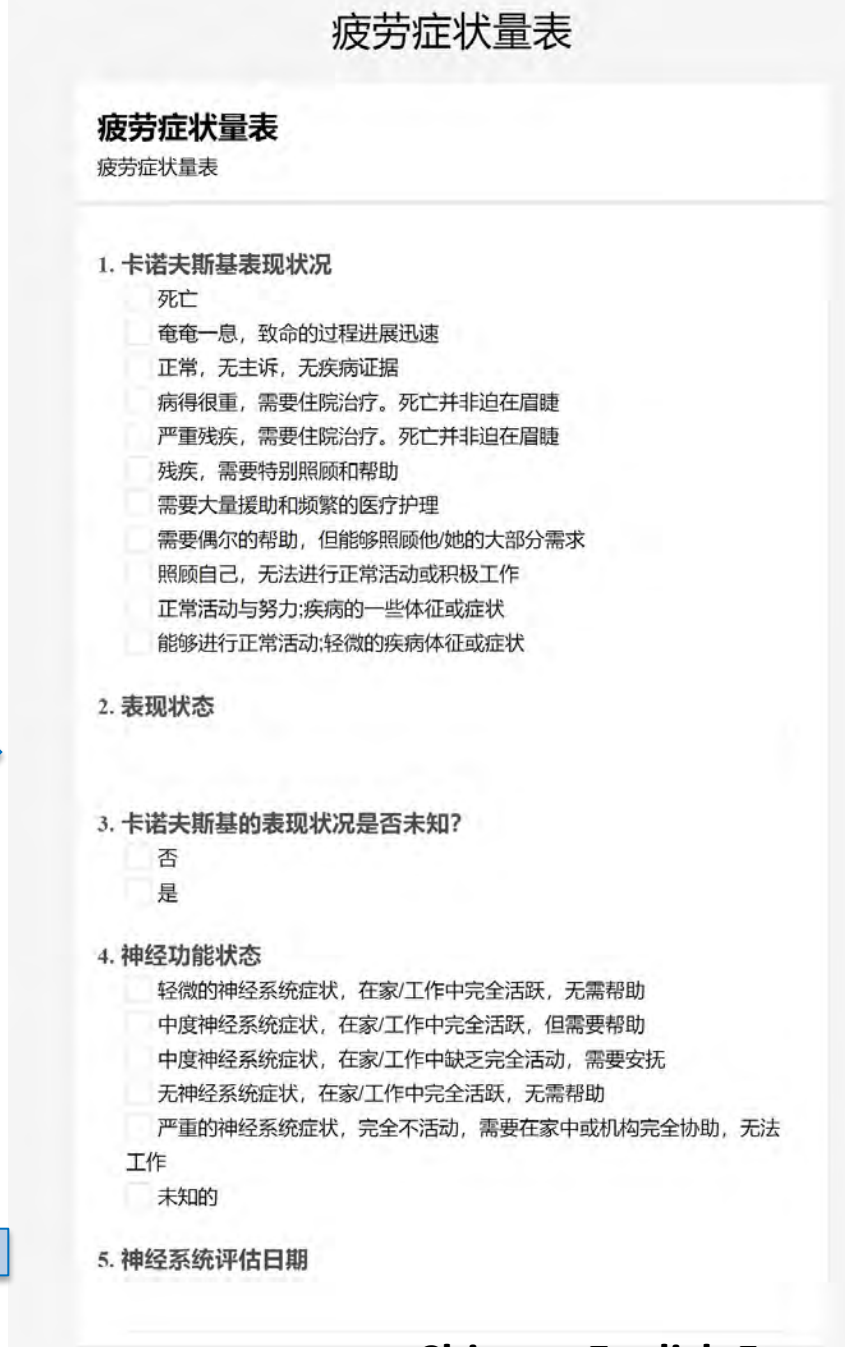


Form-Module-CDE Model

Provide precise and interoperable metadata and form for re-using

API

CDE and Form can be reused in JSON/CSV



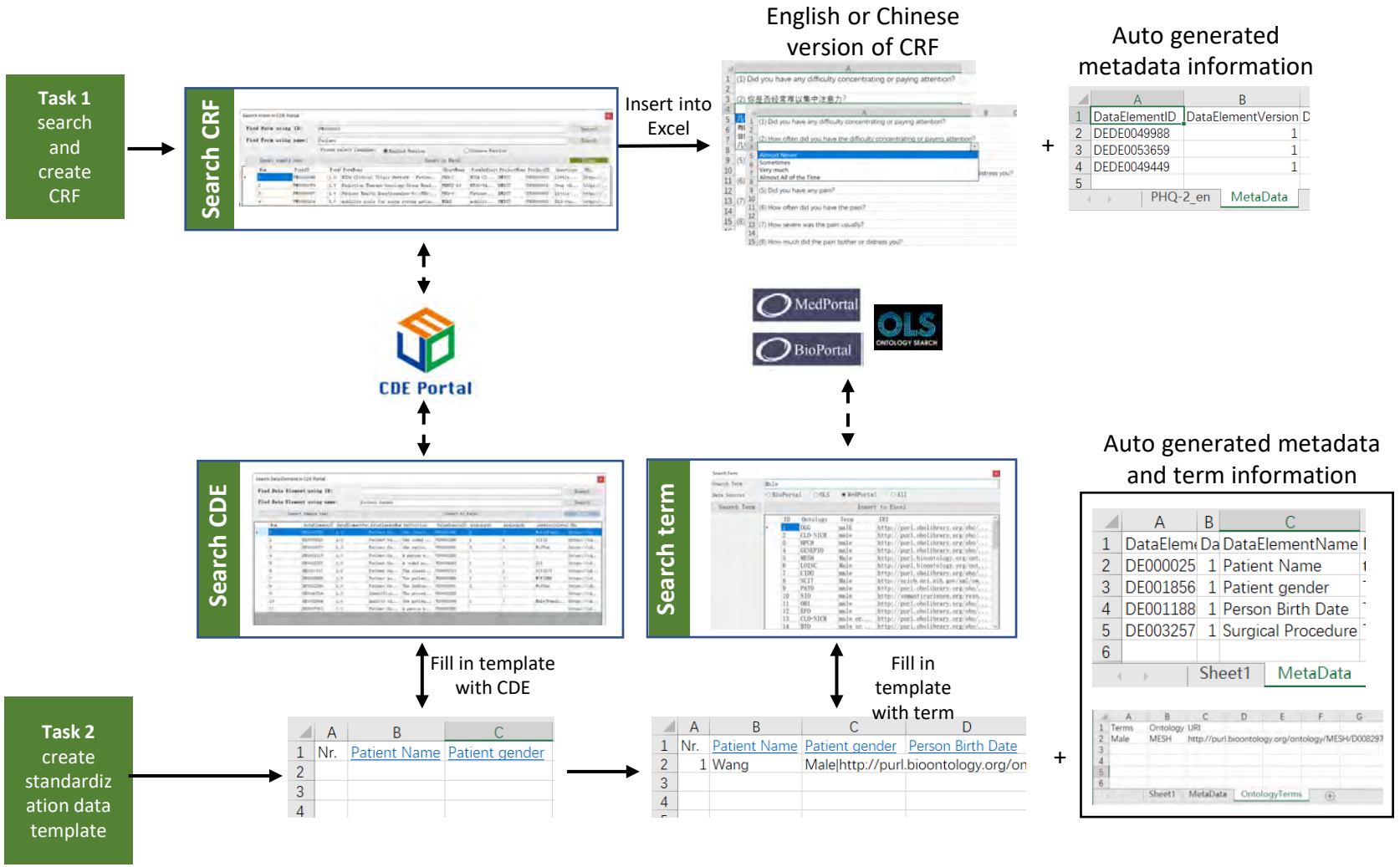
Chinese-English Form

CDE Tools: EXCEL plugin for data standard application

- **URL:**
  https://github.com/MedportalProject/CDE-Tools

- **Features:**
  - Excel plug-in
  - Ontology term search
  - Text annotations
  - Batch processing of standardized data elements

- **Extensions: Supports all ontology repositories that use OntoAlliance and OLS frameworks**



Outline

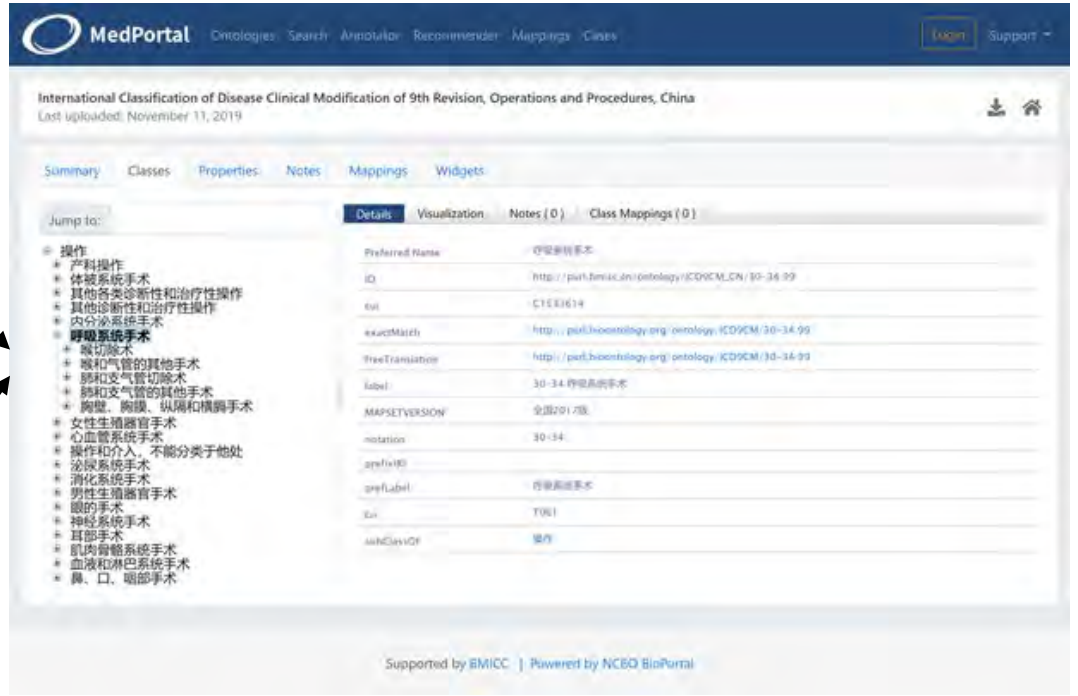
- **Goals and status quo of scientific data management**
- **Biomedical Data Standards Environment**
- **Standard resource construction**
- **Summary**

Chinese ontology resources in MedPortal

Ontology	Label	Definition	Other Annotation Properties
OGMS	76	70	4
PATO	1742	1722	906
IAO	194	183	116
OBI	1875	1828	422
BFO	36	17	170
RO	475	0	3
ICO	176	161	0
HP	9998	6659	1647
CLO*	878	330	439

Some important OBO domain ontologies were also translated into Chinese.

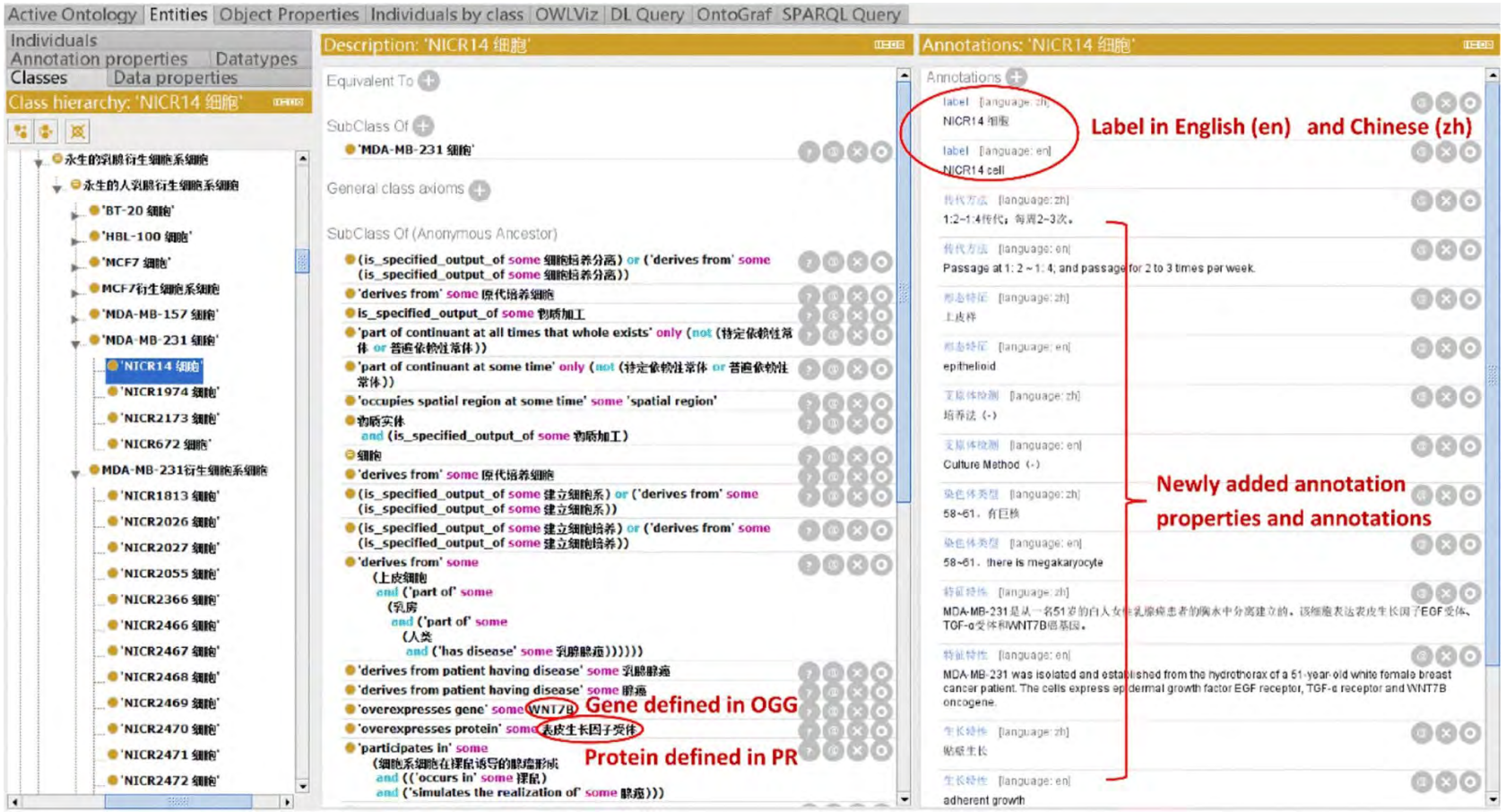
Some widely used clinical taxonomies and ontologies, such as ICD-10， ICD-11， LOINC and ICD-9-CM



Cell Line Ontology--Integrate cell line information from the Chinese experimental platform

- Chinese National Infrastructure of Cell Line Resource (NICR, 国家实验细胞资源共享服务平台)

- NICR 2 704 cell lines were integrated
- Added cell line attributes
- Formal representation of multiple biological characteristics of a cell line
- Mapping to cellular resources such as ATCC, Coriell Cell, and JCRB
- bilingual ontology subset

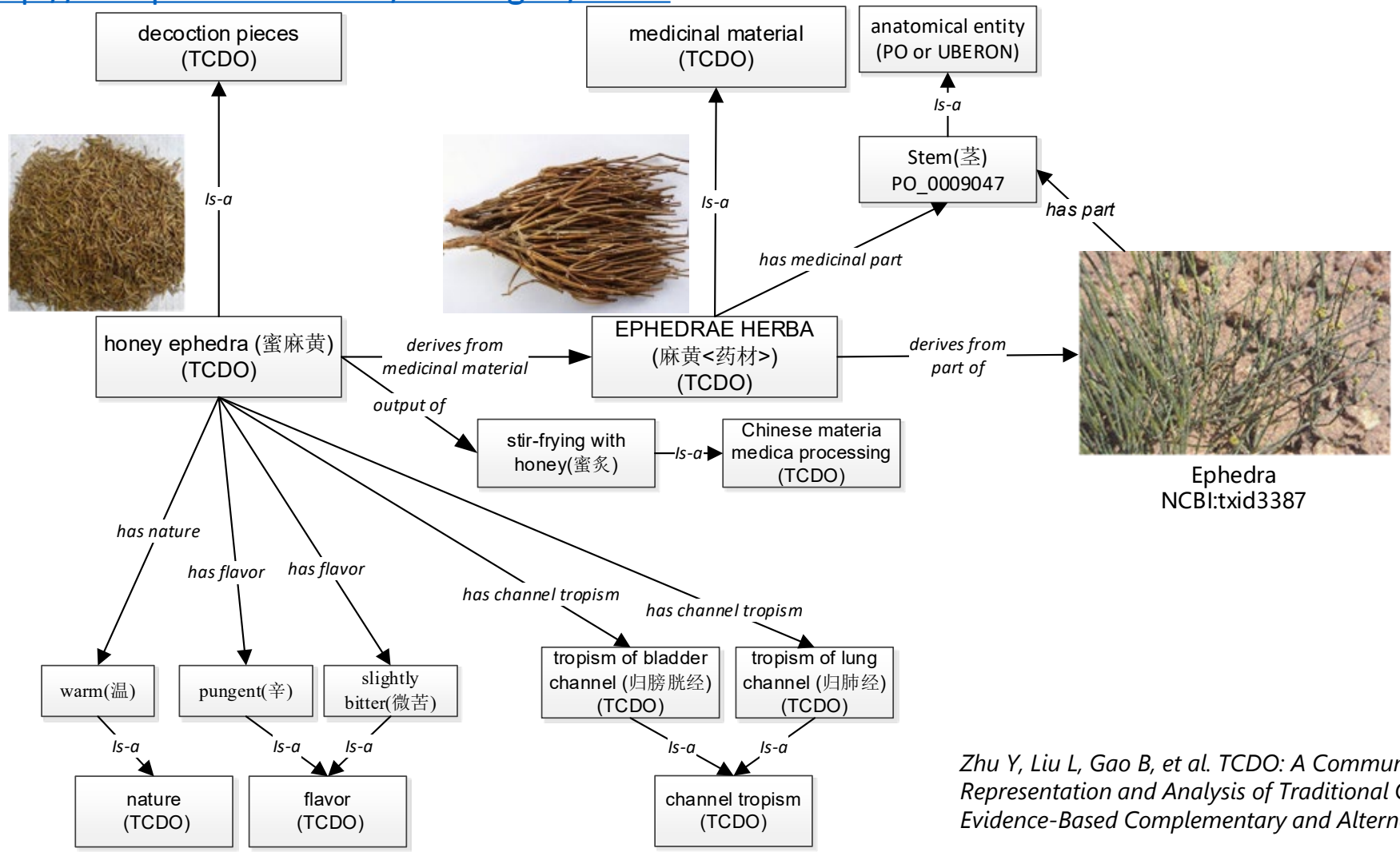


http://medportal.bmicc.cn/ontologies/CLO-NICR/

BMC Bioinformatics 20 (S5), 179 (2019)

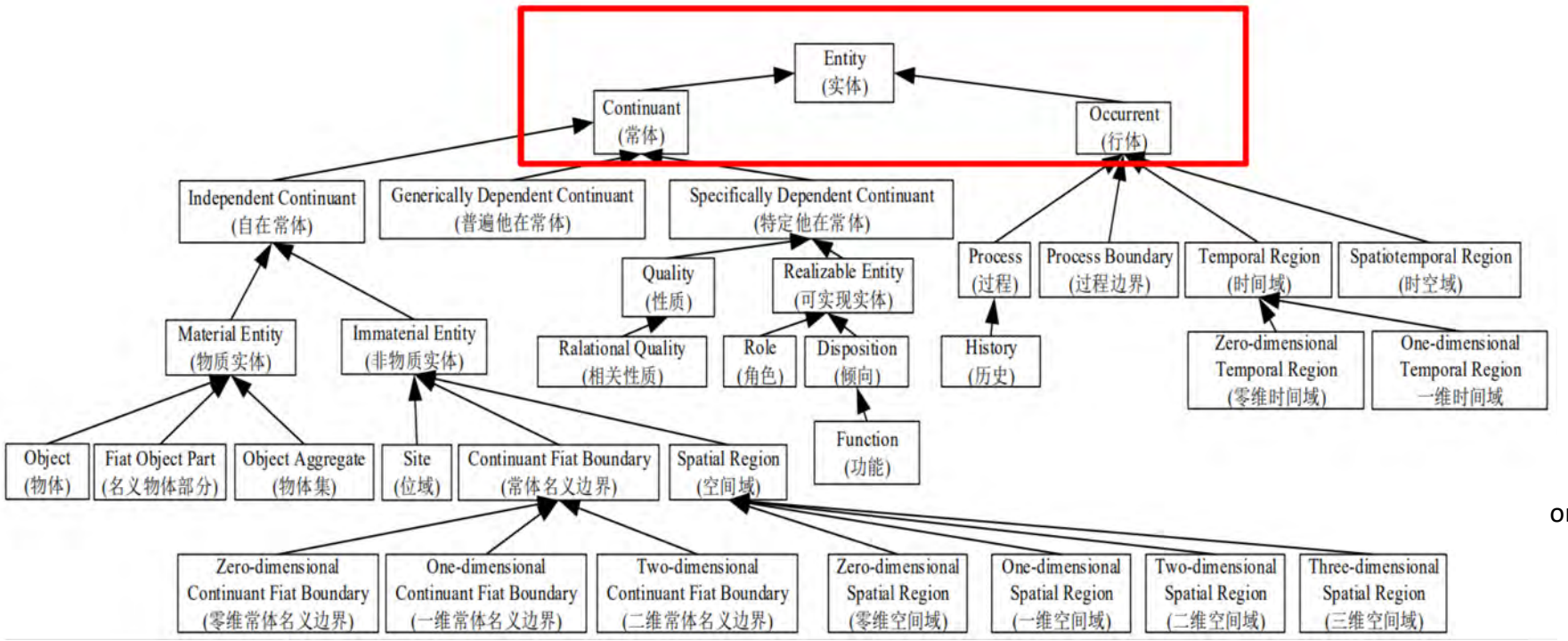
TCDO（Traditional Chinese Drug Ontology）

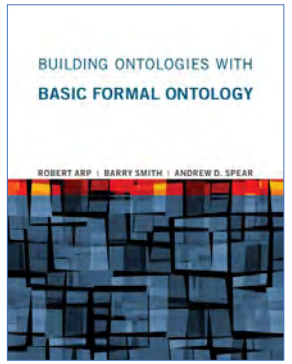
- Based BFO, more than 400 popular Chinese herbal decoction pieces and Chinese herbal medicines and their attributes are formally represented.
- http://medportal.bmicc.cn/ontologies/TCDO



Zhu Y, Liu L, Gao B, et al. TCDO: A Community-Based Ontology for Integrative Representation and Analysis of Traditional Chinese Drugs and Their Properties [J]. Evidence-Based Complementary and Alternative Medicine, 2021, 2021: 6637810.

Translate BFO (Basic Formal Ontology) into Chinese



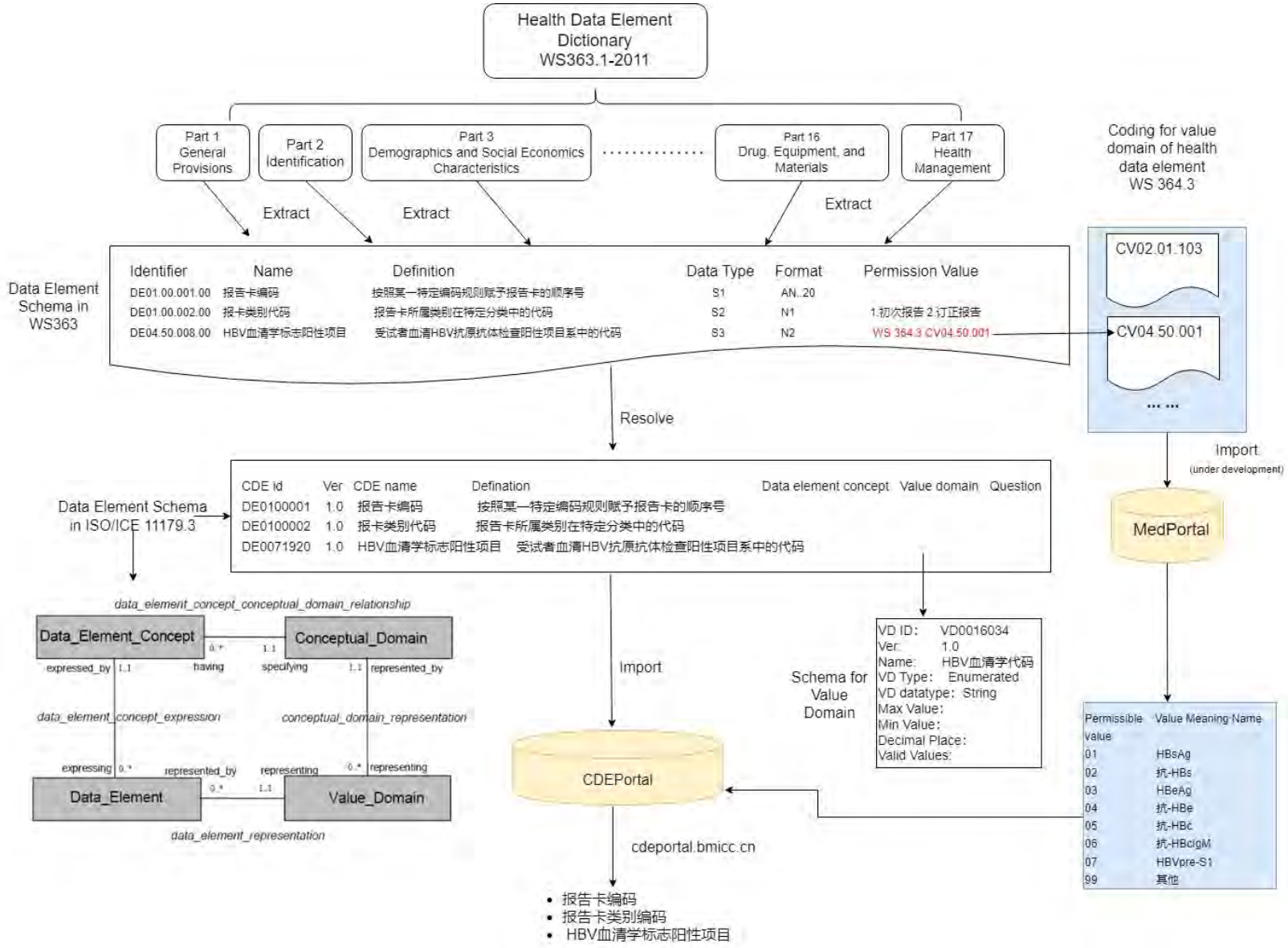


original author： Robert Arp, Barry Smith

https://medportal.bmicc.cn/ontologies/BFO-ZH

Health Data Element Dictionary （WS363.1-17&WS 364.1-17）

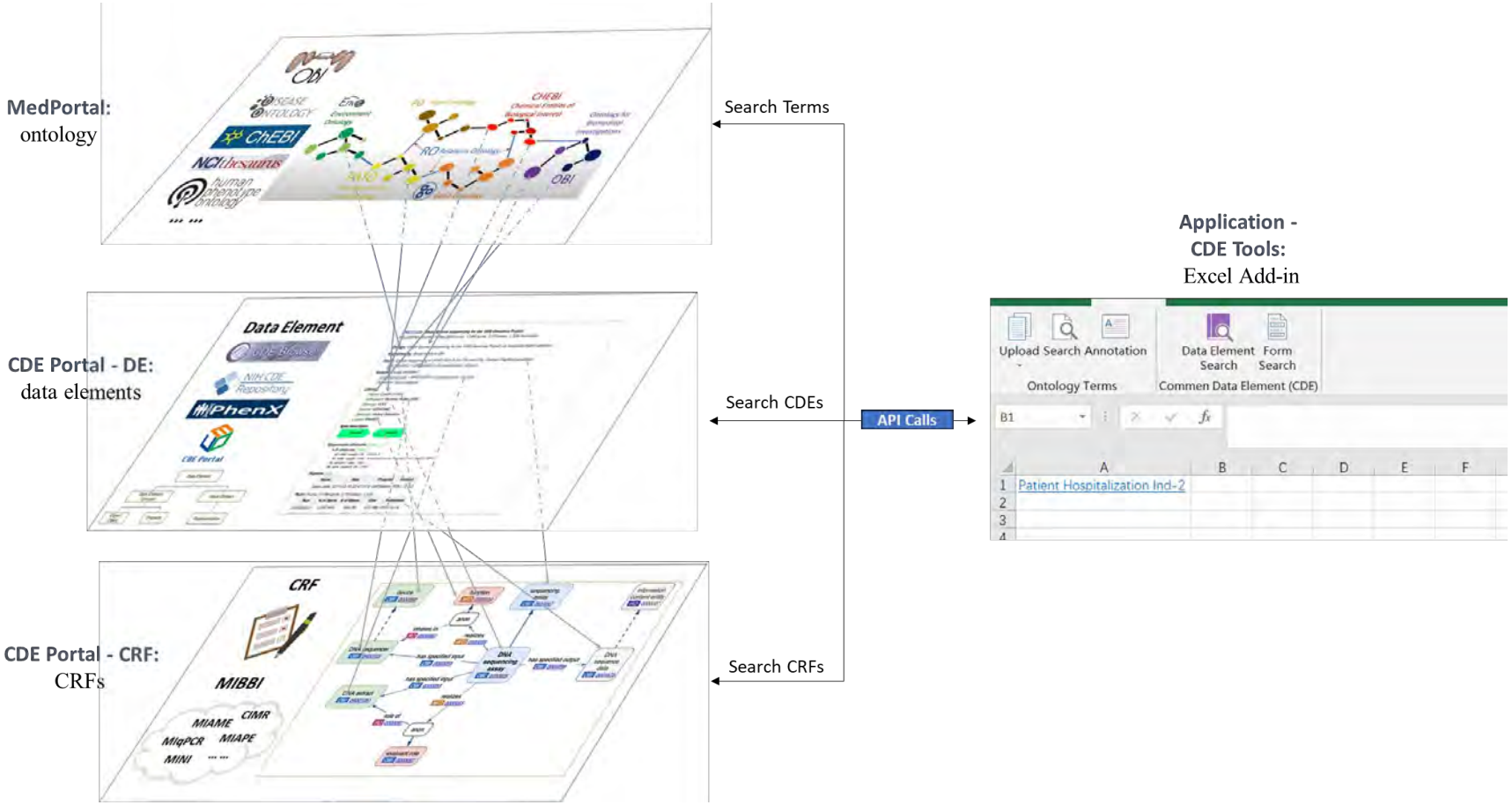
- Representation of the standards following FAIR guidelines
  - Data elements were resolved and extracted according to ISO11179 and stored in CDE Portal
  - Control vocabulary in the value domain was stored in a value field



Outline

- **Goals and status quo of scientific data management**
- **Biomedical Data Standards Environment**
- **Standard resource construction**
- **Summary**

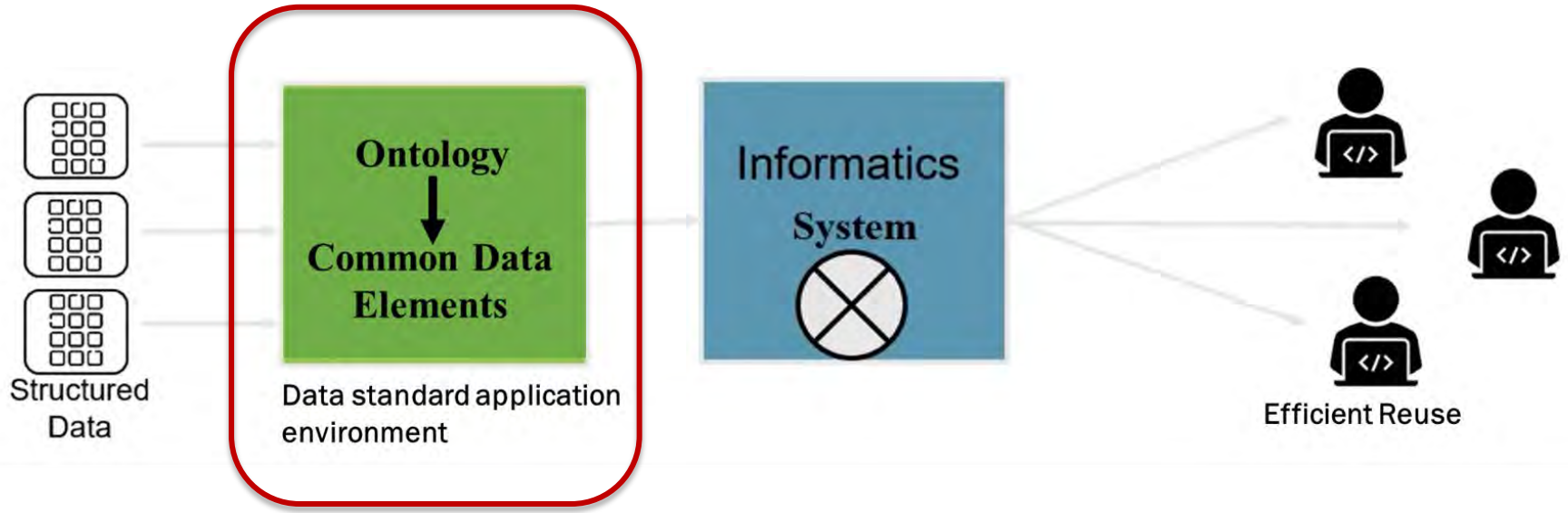
Data Types	Numbers
Ontologies	**60**
Terms	3,583,866
CDEs	**72,619**
CRFs	**106**
Value Domain	**16,711**



Biomedical Data Standards Environment

Next

- Provide support for domain-specific database in PHDA
- Build implementable biomedical data standards by using of MedPortal and CDE Portal



Acknowledgement

- Wang Heng
- Yan Zhu ((IITCM of CACMS)
- Sheng Yang
- Chen Shao
- Hongjie Pan
- Lulu Zhang
- Zhigang Wang
- Zhe Wang
- Yize Yuan
- Jingwen Guo
- Wei Zhou

- Weimin Zhu
- Yongqun Oliver He (Michigan Univ.)
- Jie Zheng(Penns. Univ.)

Thank you!