Ontological Model and Approaches in the Integrated Biomedical Database Project (IBMD)

Hiroshi Tanaka
Tokyo Medical and Dental University

What is the Integrated Biomedical DB project?

Integrated Biomedical DB Projects

- Government-commissioned project
  - Started at 2007 in **Tokyo Medical & Dental Univ.**
- by MEXT (Ministry of Education, Culture, Sports, Science and Technology)
- A subproject of the National Projects for DBs integration in Life science fields
- Main goal of this project is

  integration of disease databases in Japan

Biomedical Ontology Projects in Japan

- Two government-commissioned medical ontology projects in Japan
- Ours: Scientific DB integration by MEXT
- Clinical Information System by MHLW
  (Ministry of Health,Labour and Welfare)
  "Japan Medical Ontology Development Project for Advanced Clinical Information System"
  - **Dr. Imai's talk**
- Good collaboration between both projects

National Project for DB integration in Life Science Field (MEXT)

- Integrated DB Project in Life Science
  - **Started from 2006**

Background

- DBs in LS in Japan were scattered in various institutes
  - **DNA sequence (DDBJ) in NIG in Mishima**
  - **Protein DB (PDBJ) in Osaka University**
  - **Pathway DB (KEGG) in Kyoto University**
  - **Unlike NCBI in US and EBI in Europe**

Goal

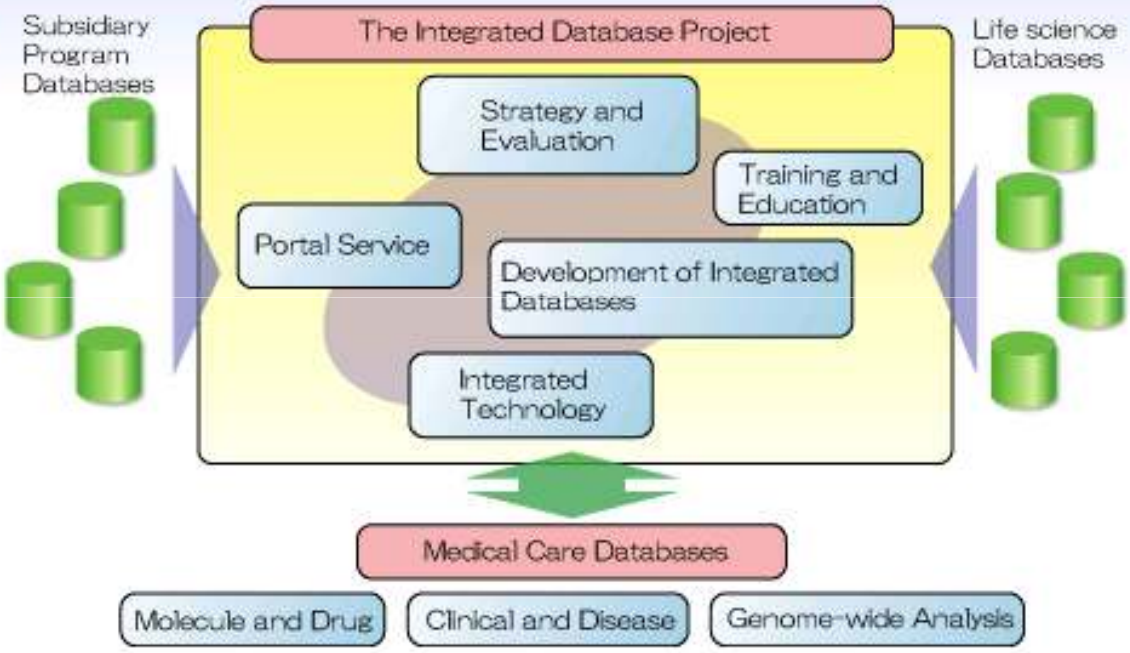
- Establish National Center for DB integration in LS (DBCLS)
  - **develop integrated DB service**
  - **common portals for DBs in LS**

Site map Japanese

search

Ministry of Education, Culture, Sports, Science, and Technology (MEXT)
Integrated Database Project

Project Overview



detail

Latest News

RSS

2009.3.31

The portal site for the "Integrated Database Project" is now available in English.

- Home
- About the Integrated Database Project
- Organization
- Achievements
- Glossary

Links for research organizations

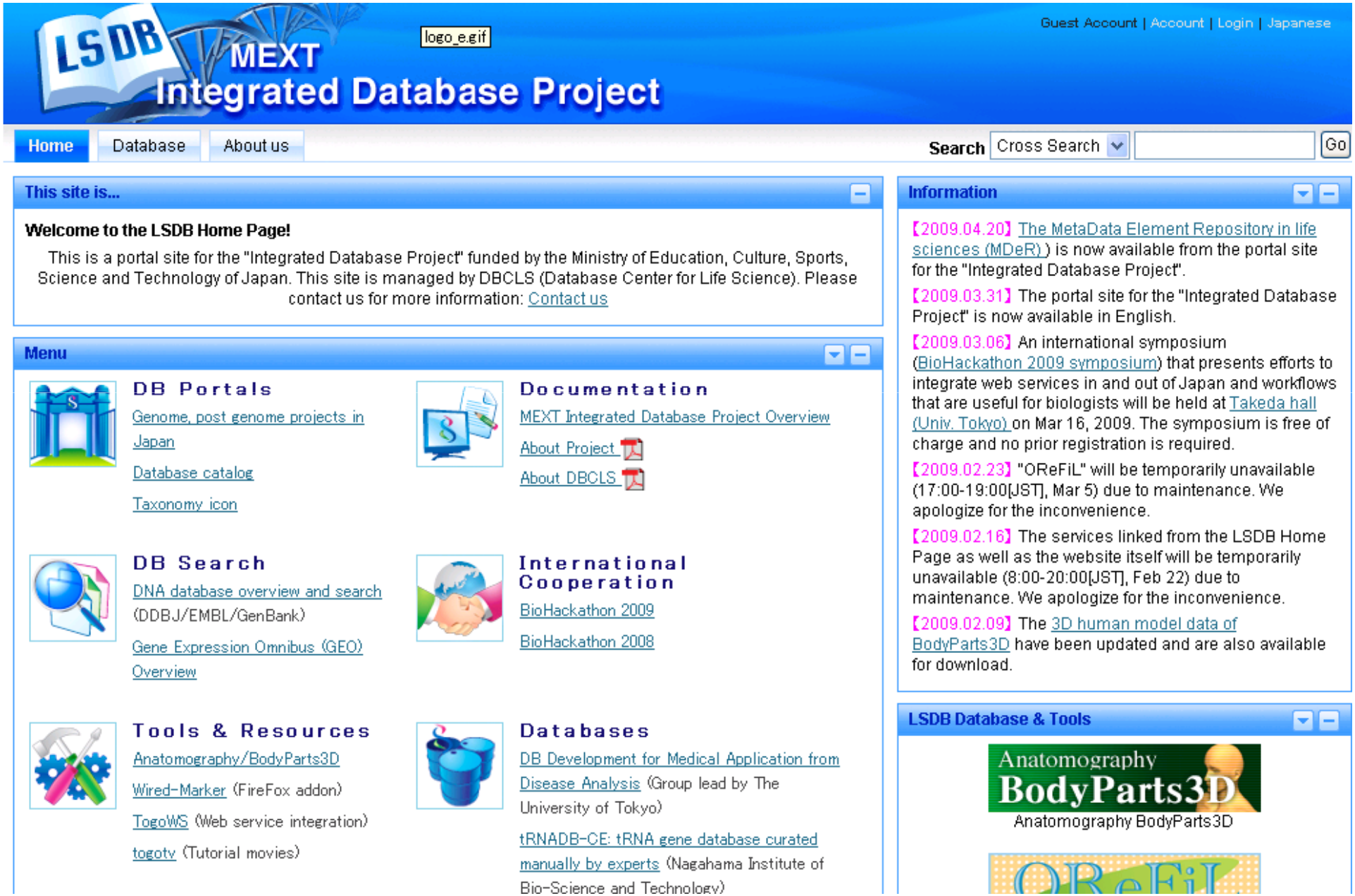
Central Organization

- Research Organization of Information and Systems (ROIS) Database Center for Life Science (Japanese version only)

Allocation organizations

- Kyoto University
- Tokyo Medical and Dental University (Japanese version only)

Integrated Database Project Portals



Genome, post genome projects in Japan

Project Name	Implementing Ministry/Agency	Start Year	End Year	Classification of Database Public Release Status	Public Database	Data Download Site
Standard SNPs Analysis Project	Ministry of Education, Culture, Sports, Science and Technology (MEXT)	1999	2001	Shared	JSNP: http://lifesciencedb.jp/?pg=1&tp_dbsid=285	SNP frequency data, etc. (ftp://ftp.hgc.jp/pub/hgc/db/snp/) XML (mapping data, etc.) (http://snp.ims.u-tokyo.ac.jp/XML.html) Search tool (http://snp.ims.u-tokyo.ac.jp/map/Dump/)
Genome Diversity Project	Ministry of Economy, Trade and Industry (METI)	2000	2005	Shared	-	- GDBS (http://jbirc.jbic.or.jp/gdbs/database/viewer/download/ - The Integrated Cancer Genome Database. (http://genomecenter.jfcr.or.jp/genomedb/)
Standardization Survey of Data Description Format for Genotype-Phenotype Database	Ministry of Economy, Trade and Industry (METI)	2006	2008	-	-	-
Integrated Database Project	Ministry of Economy, Trade and Industry (METI)	2000	2004	Shared	H-InvDB: http://lifesciencedb.jp/?pg=1&tp_dbsid=491	http://hinvdb.ddbj.nig.ac.jp/ahg-db/download.jsp
Genome	Ministry of					

Allocation Organization for Integrated DB Projects

- Tokyo Medical and Dental University (TMDU) , Information Center for Medical Science
  - commissioned by MEXT
  - to integrate scattered databases on various disease in Japan
  - to be a National Center for Disease Database Integration
  - from 2007, Integrated Biomedical (clinical) DB
- Kyoto University Bioinformatics Center
  - Integration of Drug Information
- Tokyo University Dept. Human Genetics
  - Integration of Genetic Polymorphism

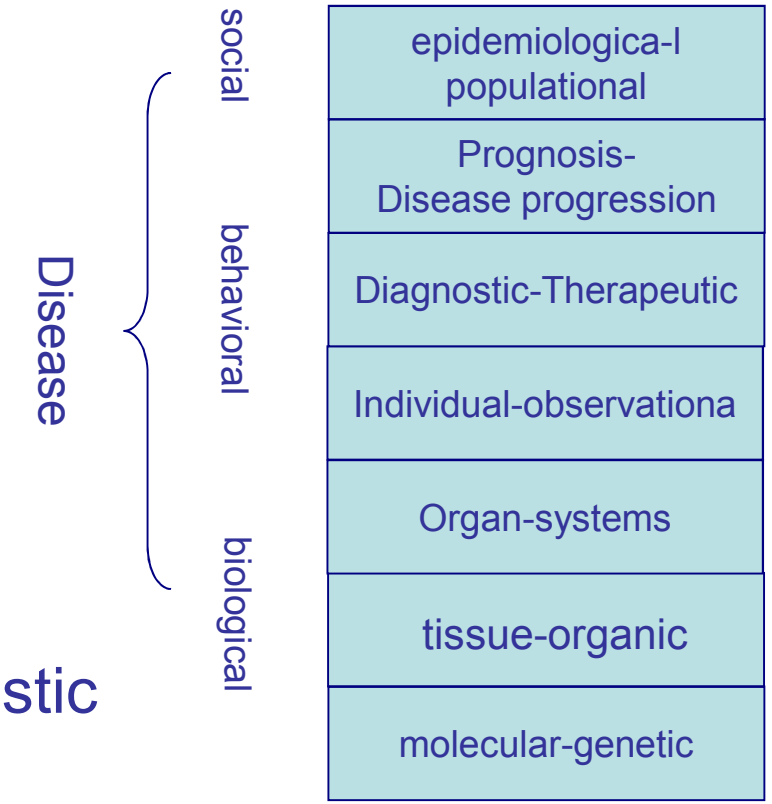
Challenges for Integrated Biomedical DB projects

- Development of Ontology-based Integrative Interface for Disease DB search
- Integration of the Core Clinical DBs in Japan
  - Parkinson Disease DB in Osaka Univ.
  - GEMDBJ in National Cancer Center
  - So on
- Establish Ethical Code for Publicizing Clinical Case DB

Then, how can we describe
Disease?
Is it physical entity, or
just conceptual ?

Disease Modeling in Database

- What is the ontological characteristics of disease?
- Complexity of “Disease”
- Disease is a multifaceted, multilayered entity
  - molecular-genetic
  - tissue-organic
  - individual
  - diagnostic-therapeutic-prognostic
  - behavioral (medical practice)
  - populational-epidemiological

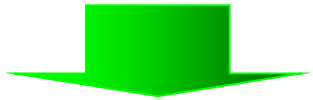


Disease Modeling in Database

- Complexity of the “Disease”
  - Multi-faceted, multi-layered entity
  - Incompleteness of disease knowledge
  - Multiplicity of related sciences
    - biological, psycological, behavioral, conceptual, social science
- Main opposite standpoints
  - Physical vs Conceptual
  - Causative vs Observational

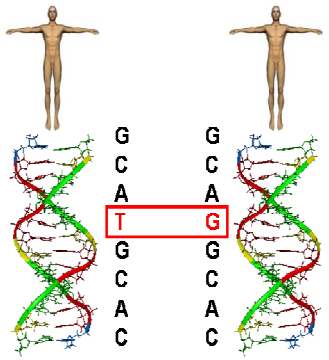
How do we describe disease

- Formal Description of Disease
- Disease View are now changing since Revolution of Molecular medicine
- Conventional View
  - Multilayered Phenotypical Description
  - Essentially Observational
  - (Place, Organ) X (Pathomorphology)
    - myocardial (place) infarction (pathology)

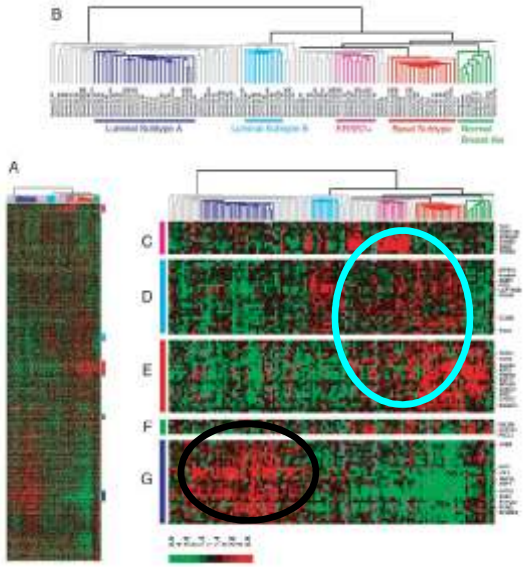


Recent Changes of Disease View

- Advances in Molecular Medicine
  - – disease genetics
    - disease causative (related) gene
    - genetic polymorphism (SNP, ms)
  - – disease omics
    - genetic expression profile
    - proteome, metabolome
  - – disease molecular pathway
    - distorted signal pathway or regulatory network

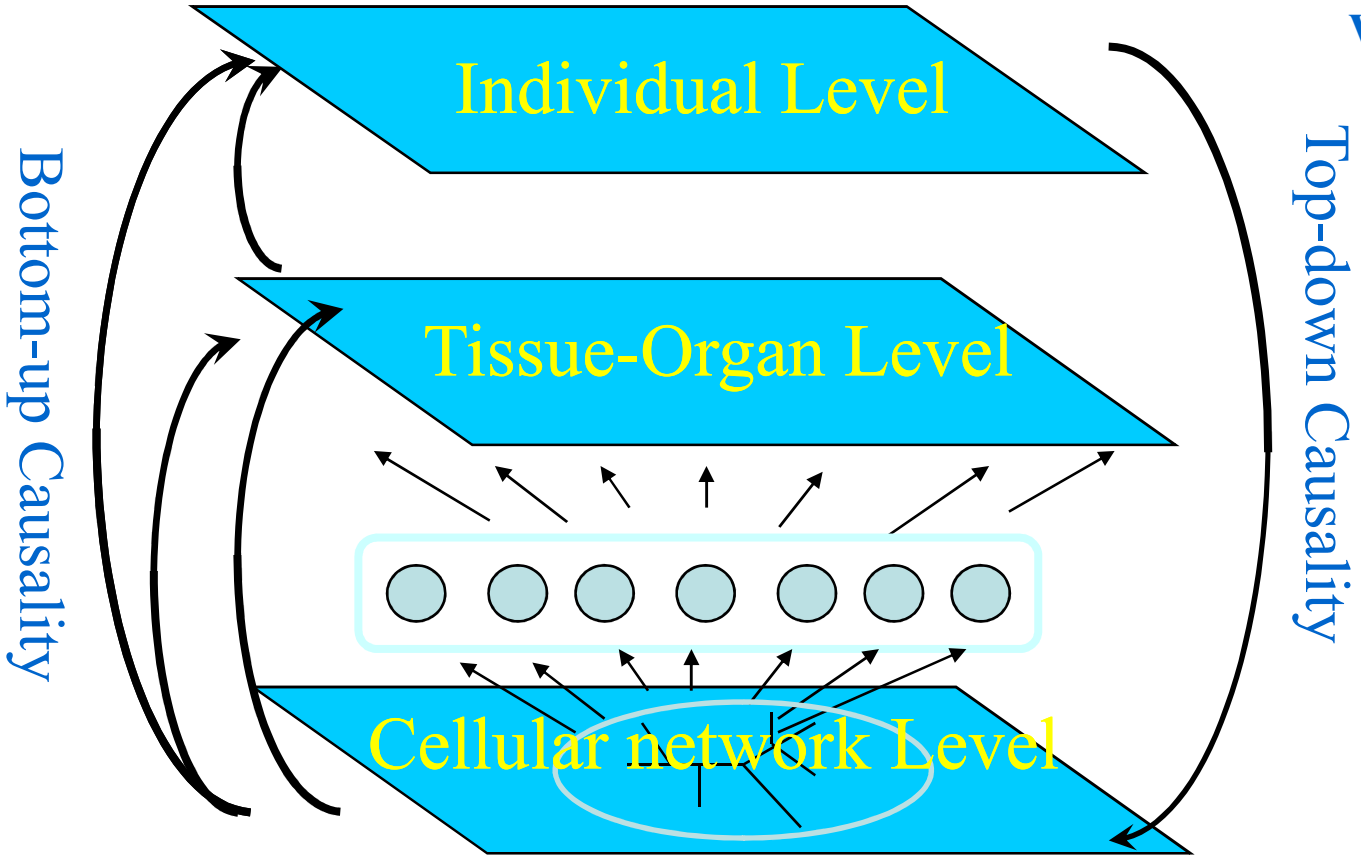


SNPs



DNA microarray

Disease is hierarchically organized
"distorted molecular network"



Whole Body Coordination

Tissue-Morphological abnormality

---

cell-cell communication

---

molecular pathway

signalling network
gene regulatory net
metabolic network

Ontology

- Formal representation of a set of concepts within a domain and the relationships between those concepts
  - used to reason about the properties of that domain, and to define the domain
  - “Formal, explicit specification of a shared conceptualisation“ (Gruber, kls stanford)
- Controlled (Formal) Vocabulary
  - used to model a domain for knowledge sharing and reuse
  - the type of objects /concepts that exist, and their properties and relations

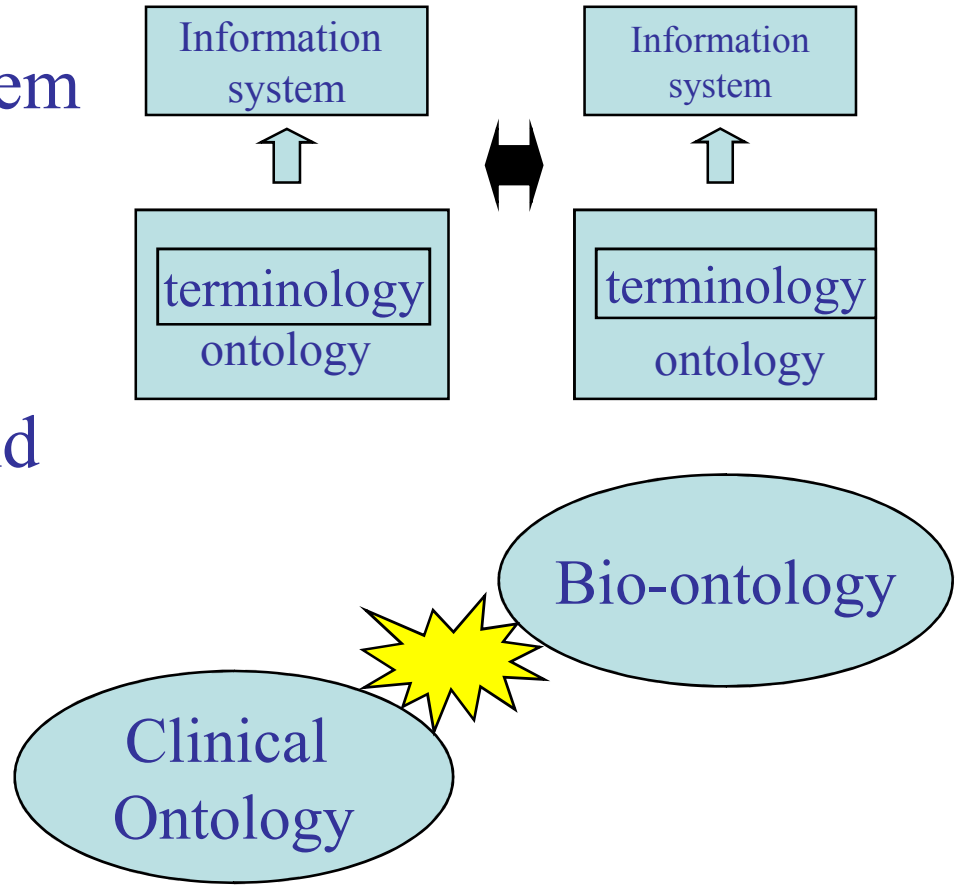
Challenges in Biomedical Ontology

Ontology mismatch between clinical thinking and Omics medicine

- Each information system has its ontology as a basis
- Mismatch between

  Clinical ontology and

  Bio-ontology

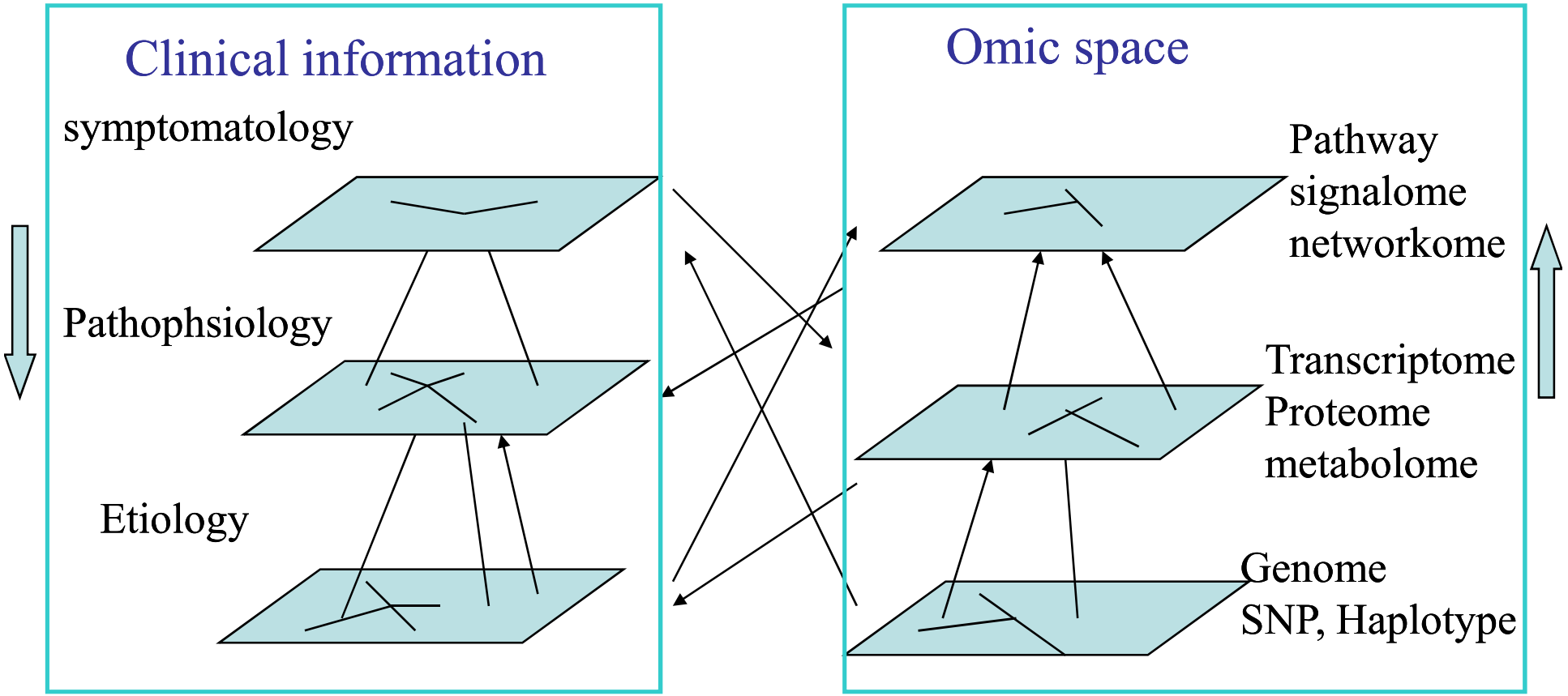


Mismatch of Thinking

- Clinical thinking
  - Organs and diseases are units of concepts
  - Clinical phenotypical; disease is defined on pathological, morphological (changes) base
  - Essentially goal-oriented: disease care directed
  - Top down
- Molecular (Omics) medicine thinking
  - Molecular function and their functional relation to other molecules
  - Products of gene expression are units of concepts
  - Bottom up

Two Worlds

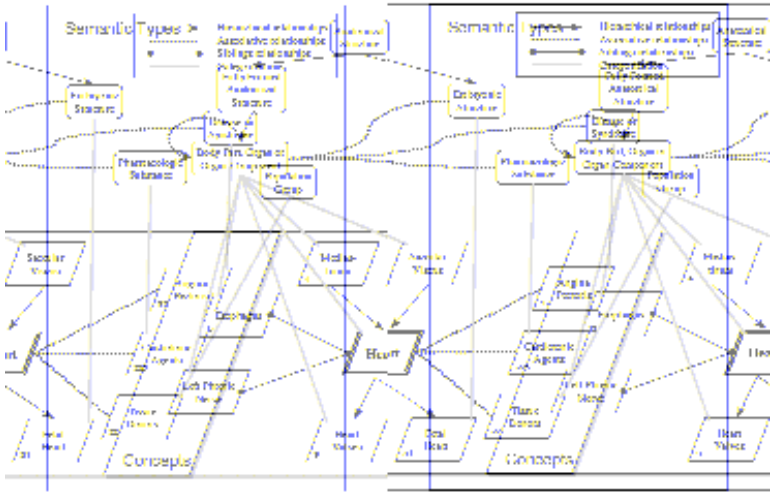
Clinical thinking and "Omic space"



Association Map

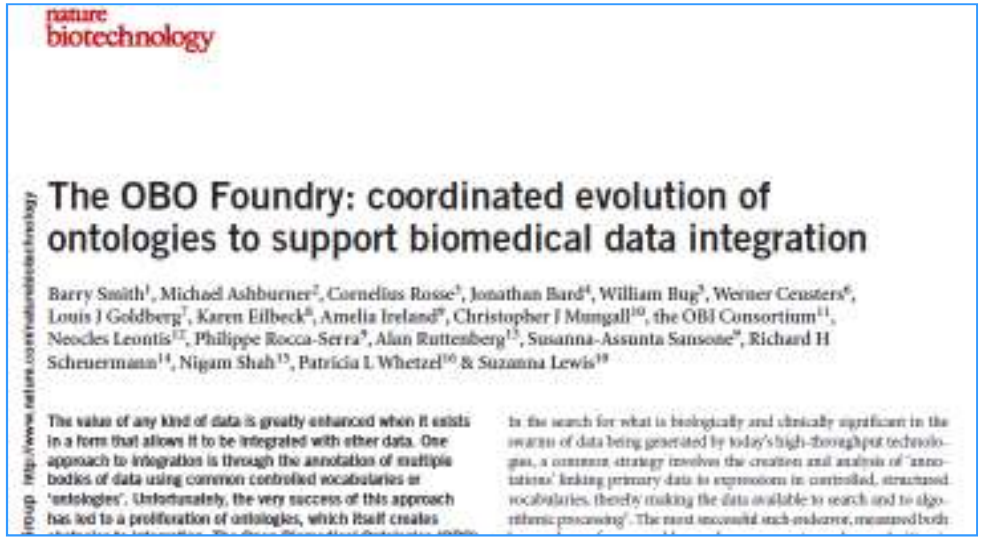
Existing Clinical and Bio-Ontology

- Clinical ontology
  - Semantic network, UMLS, Galen (SNOMED)
- Gene Ontology
  - Molecular function, process, cellular location of gene products
  - Now only, Eukaryotes



Several projects for Integration of Bio/Clinical Ontology

- Open Biomedical Ontologies (OBO)
  - National Center for Biomedical Ontology (Mussen)
  - OBO Foundry (Smith)
- Other Projects
  - Ontology for Biomedical Investigations
  - ULMS plans to involve Gene Ontology
  - Disease Ontology
  - So on



nature biotechnology

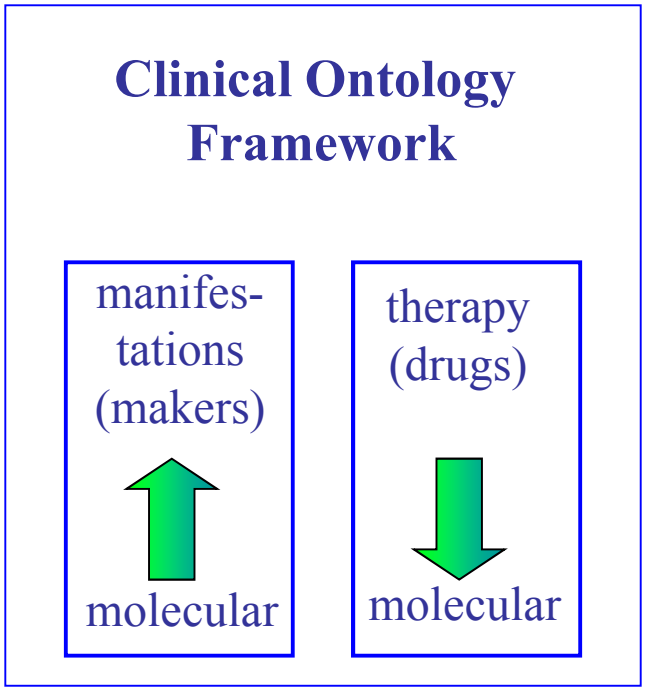
The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration

Barry Smith[1], Michael Ashburner[2], Cornelius Rosse[3], Jonathan Bard[4], William Bug[5], Werner Ceusters[6], Louis J Goldberg[7], Karen Eilbeck[8], Amelia Ireland[9], Christopher J Mungall[10], the OBI Consortium[11], Neocles Leontis[12], Philippe Rocca-Serra[9], Alan Ruttenberg[13], Susanna-Assunta Sansone[9], Richard H Scheuermann[14], Nigam Shah[15], Patricia L Whetzel[10] & Suzanna Lewis[10]

The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates

Integrative Clinico-Omic Ontology (possible transit form)

- Global structure follows the framework of clinical ontology
- Within the framework bio-ontology is employed to provide bottom-up relation of the meaning of phenotypical entities



Practical Solution

- Clinical Nosological Ontology
  - Coventional Textbook knowledge Ontology with patient medical information ontology
  - used for Ontology for Disease DBs integration
- Multilayered Clinical Omics Ontology
  - Still under development
  - But with linked multilayered data schema
  - integrated Clinical Omics Database (iCOD)

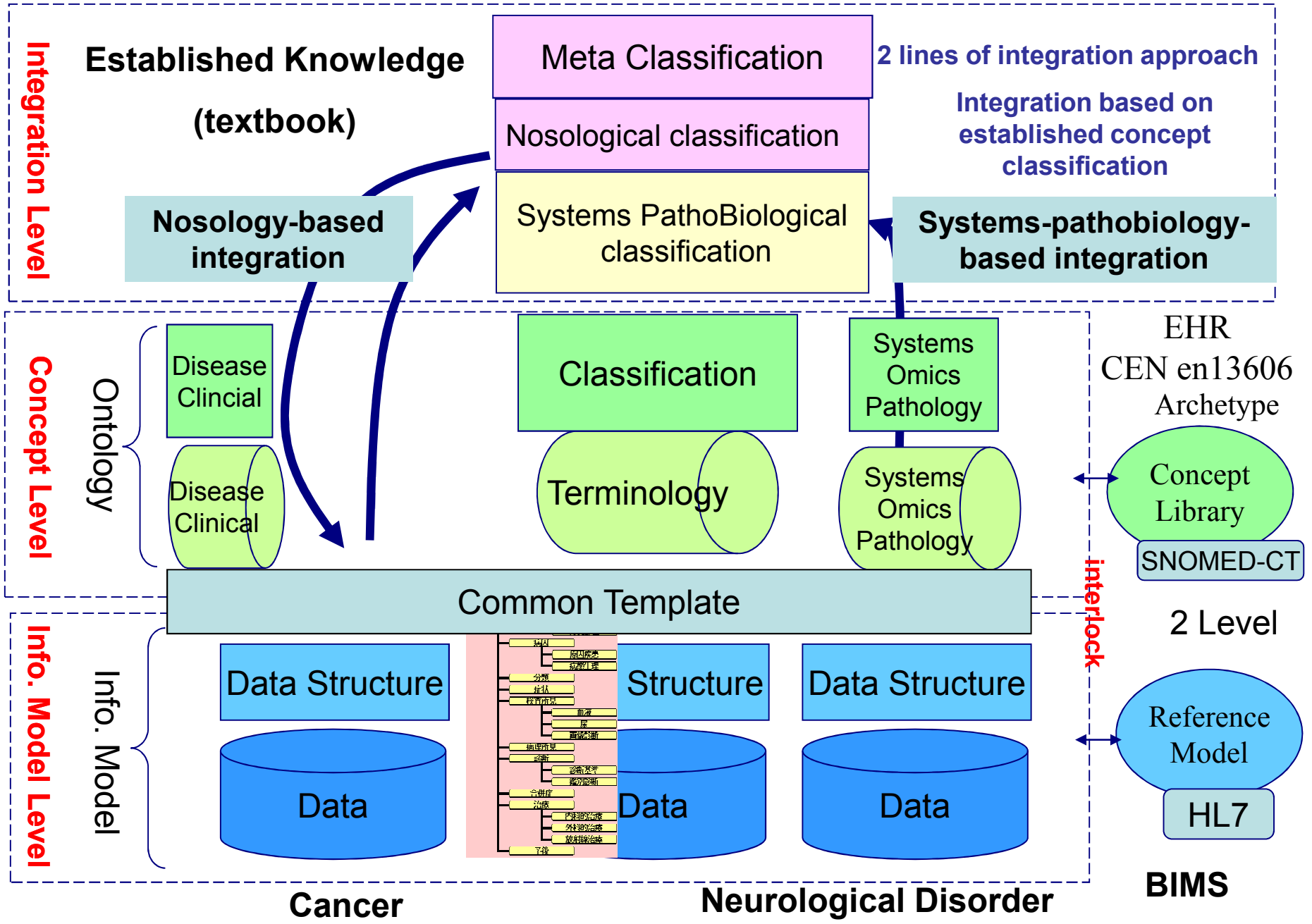
Several Results of IBMDB project

Results for Ongiong IBMDB project

- Developed the first version of Nosological IBMBD Ontology
- Mutlilayered Integrative Clinical Omics DB
- Developed Ontology-based InterDB Search System
- Trial system for integrative DB search between Parkinson Disease DB and iCOD

IBMDB Ontology

3 Level Method: 3 Level model for database integration



Disease Terminolology and Classification

Nakaya, J., Sasaki, K., and Tanaka, H. (2006) Condensed Cross Clinical Knowledge, Computer Science, IJCSNS. 6 (7A). 6-11.

Disease Classification

Condensed Cross Clinical Ontology

Disease Terminology

This is the disease terminology and classification.
Diseases are classified with combinated anatomical hierarchy and etiological hierarchy.
A conceptual unit of a Diseasse is described with the 3rd normalized skeleton template which can be called as a content model.

Anatomical Hierarchy

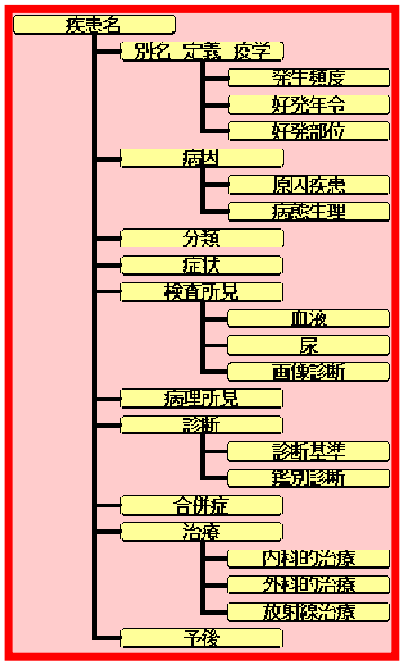
＋

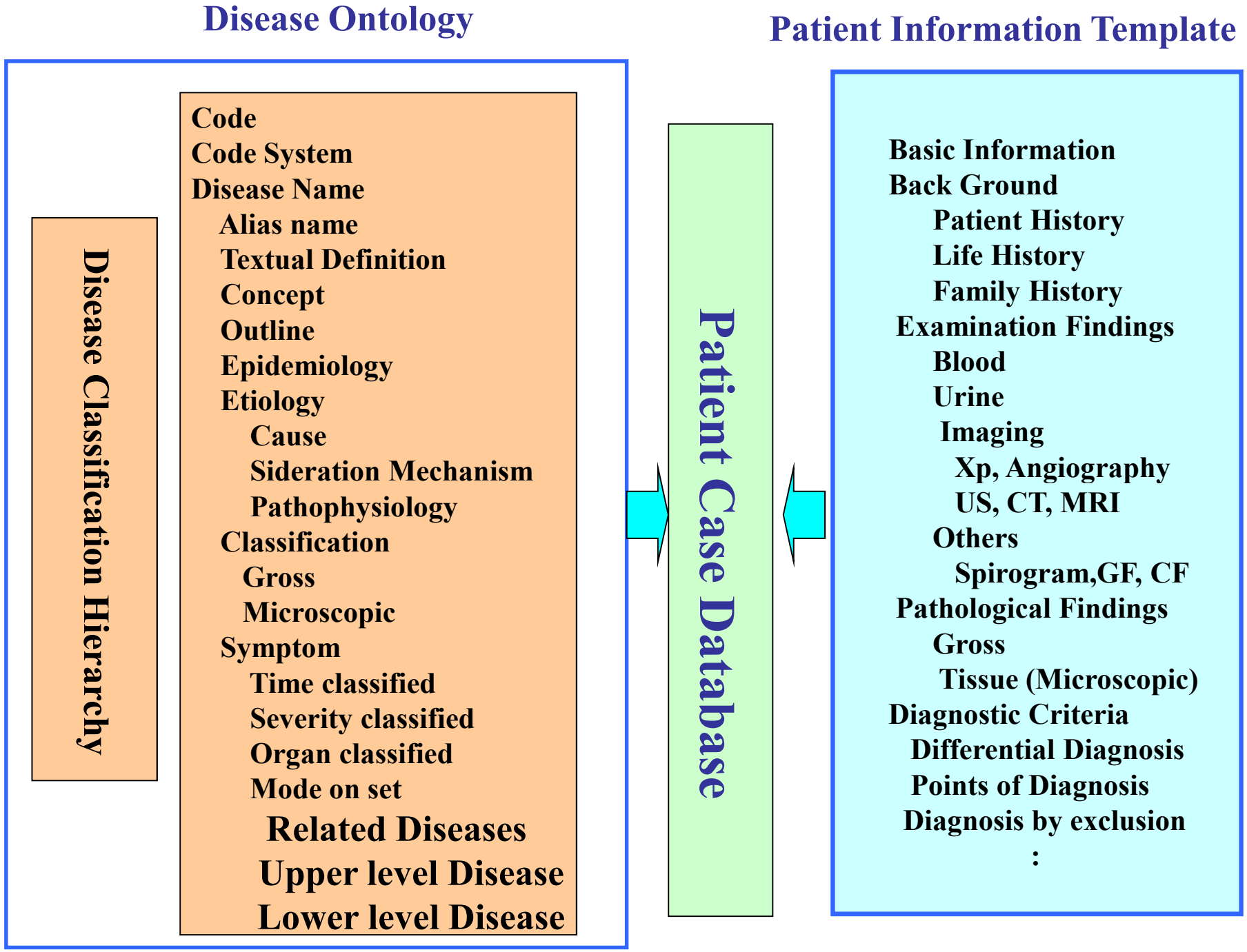
Etiological Hierarchy

Patient Information :

Normalized Skeleton Template for Disease

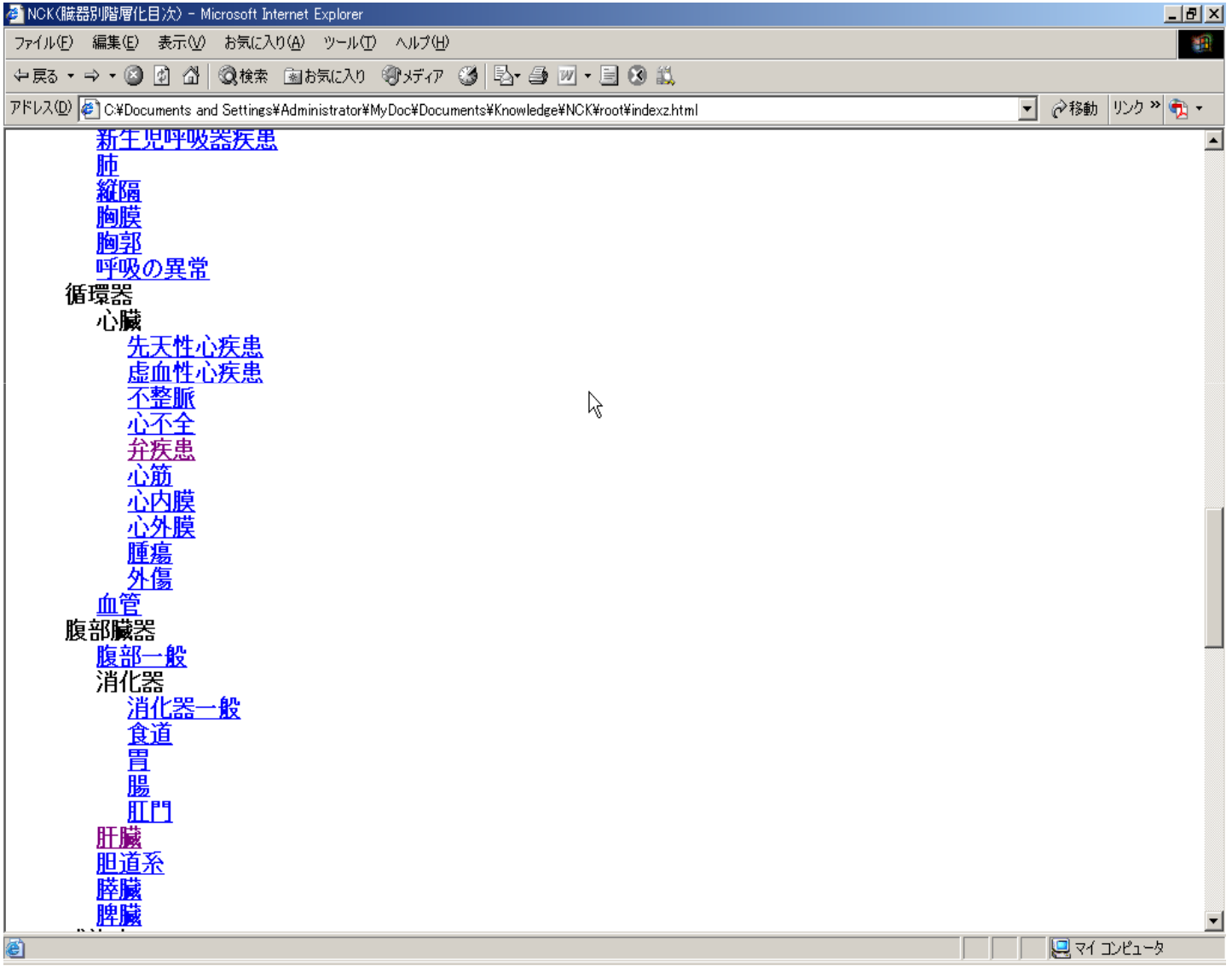
A Content Model

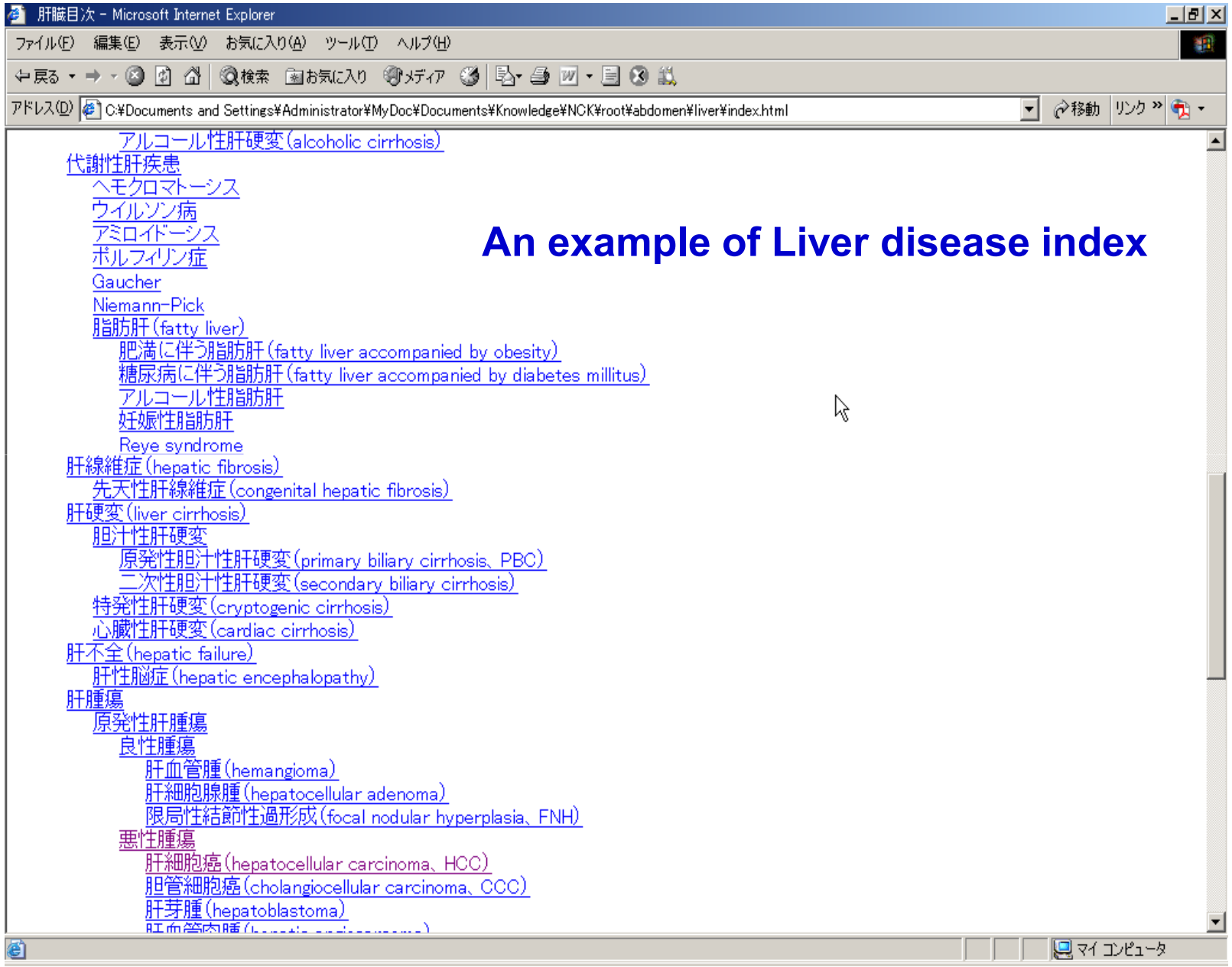



Disease Ontology
Patient Information Template
Disease Classification Hierarchy
Code
Code System
Disease Name
Alias name
Textual Definition
Concept
Outline
Epidemiology
Etiology
Cause
Sideration Mechanism
Pathophysiology
Classification
Gross
Microscopic
Symptom
Time classified
Severity classified
Organ classified
Mode on set
Related Diseases
Upper level Disease
Lower level Disease
Patient Case Database
Basic Information
Back Ground
Patient History
Life History
Family History
Examination Findings
Blood
Urine
Imaging
Xp, Angiography
US, CT, MRI
Others
Spirogram,GF, CF
Pathological Findings
Gross
Tissue (Microscopic)
Diagnostic Criteria
Differential Diagnosis
Points of Diagnosis
Diagnosis by exclusion
:

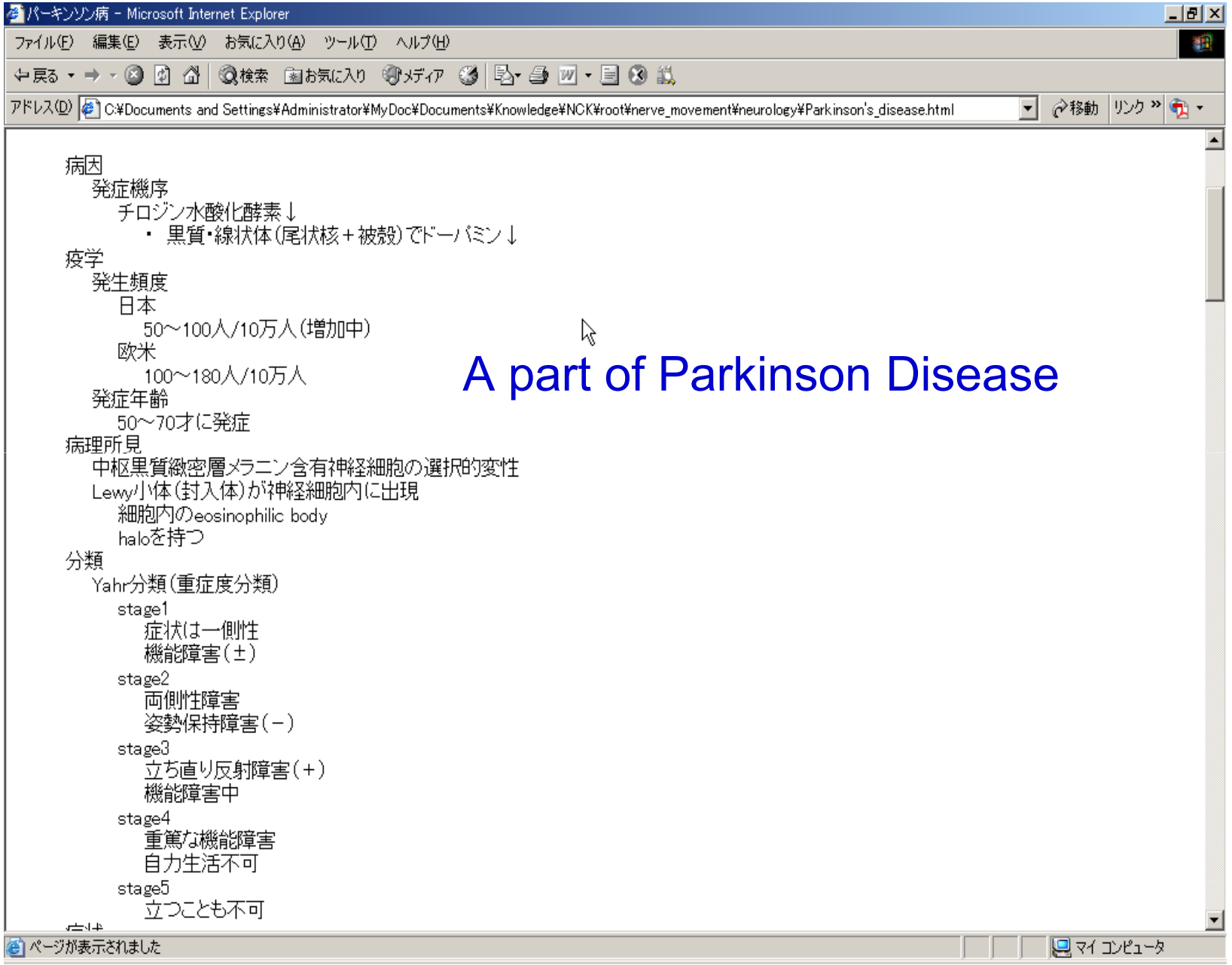
Nosological Ontology

（an example of Composite index of anatomy and etiology)





An example of Liver disease index

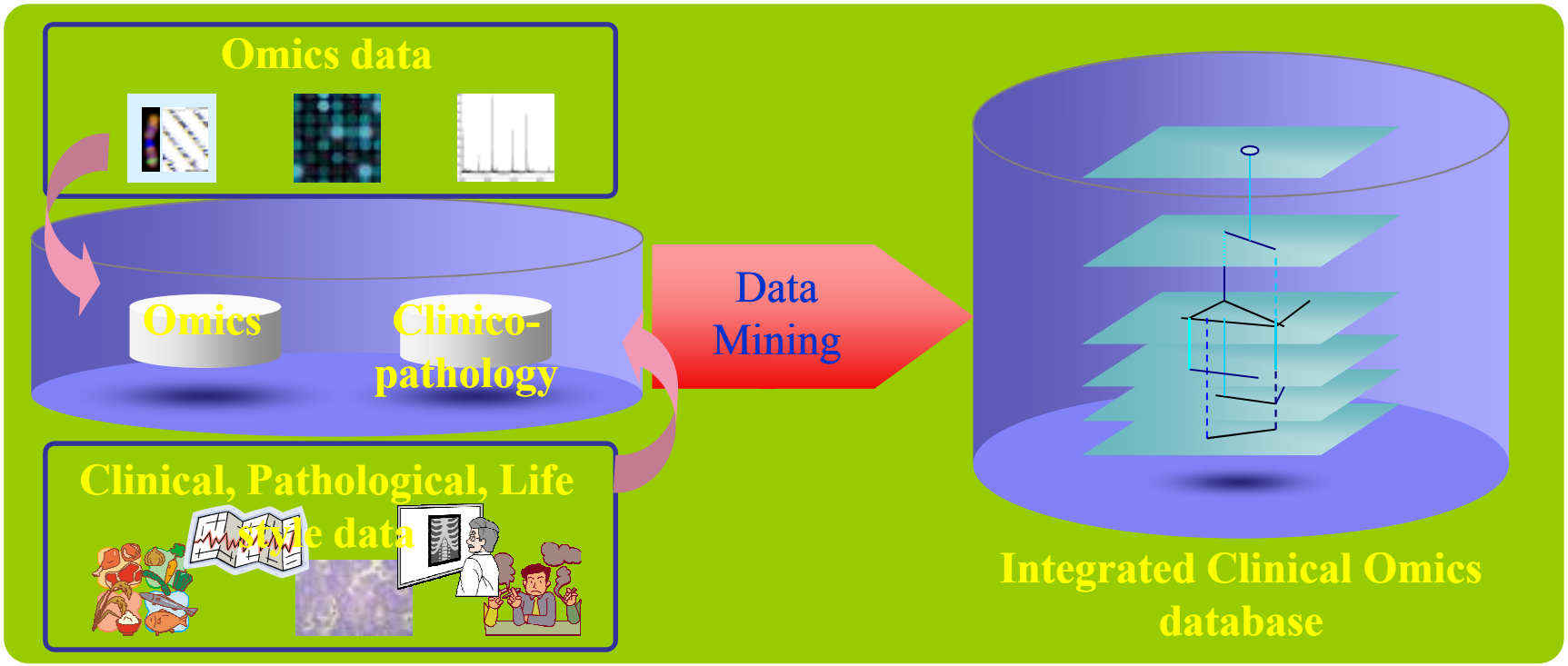


A part of Parkinson Disease

integrated Clinial Omics DB (iCOD)

Purpose

- Integrate the molecular omics information and clinical and pathodological, life style information
- Comprehensive database based on the concept of "omics-based systems medicine"



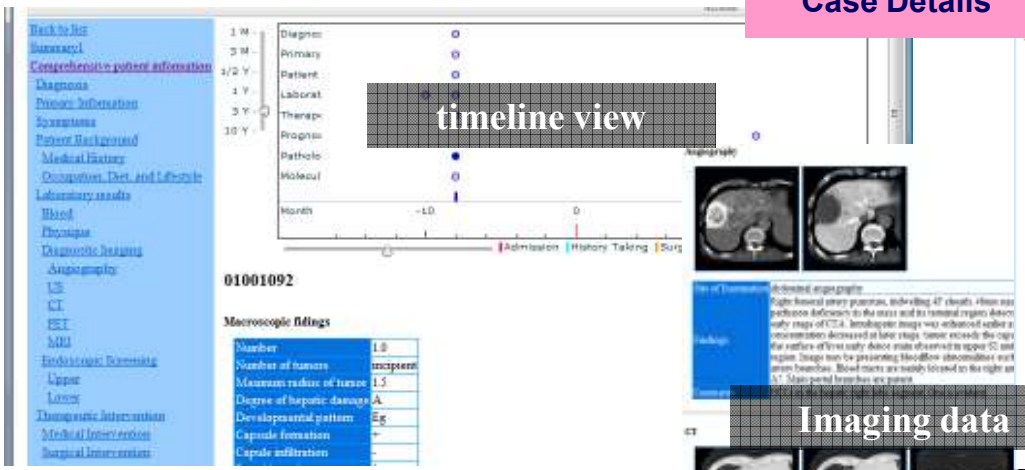
Cases

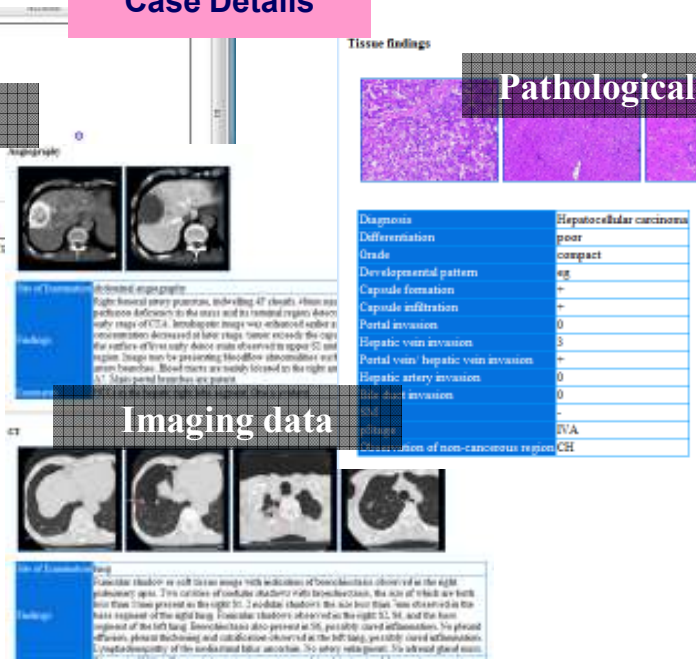
		cases	omics information			
			transcriptome		CNV	
			specimen	normal	specimen	ormal
Hepatocellular carcinome		**193**	**152**	**96**	**102**	**35**
	stored	41	29	0	34	0
	fresh	134	105	81	66	35
	metastasis	18	18	15	0	0
colon cancer		**184**	**131**	**28**	**39**	**40**
	colon	128	102	28	36	36
	rect	37	29	0	3	4
oral tumor		**148**	**20**	**0**	**64**	**2**
	stored	64	0	0	64	2
	fresh	84	20	0	0	0
total		**525**	**303**	**124**	**205**	**77**

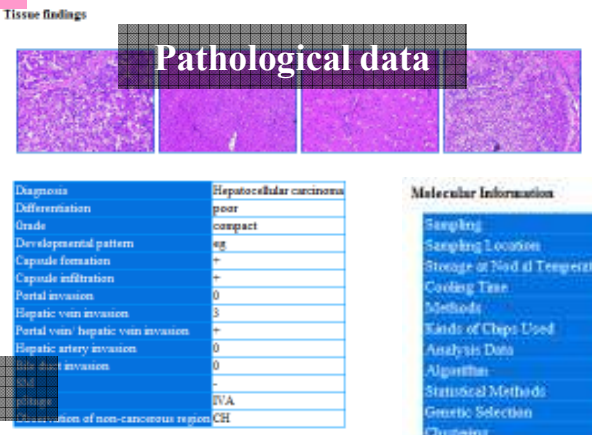
iCOD: integrated Clinical Omics Database

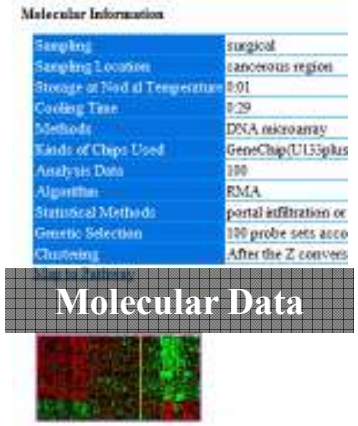
Top page → Case Archvie → Case List

Case Details









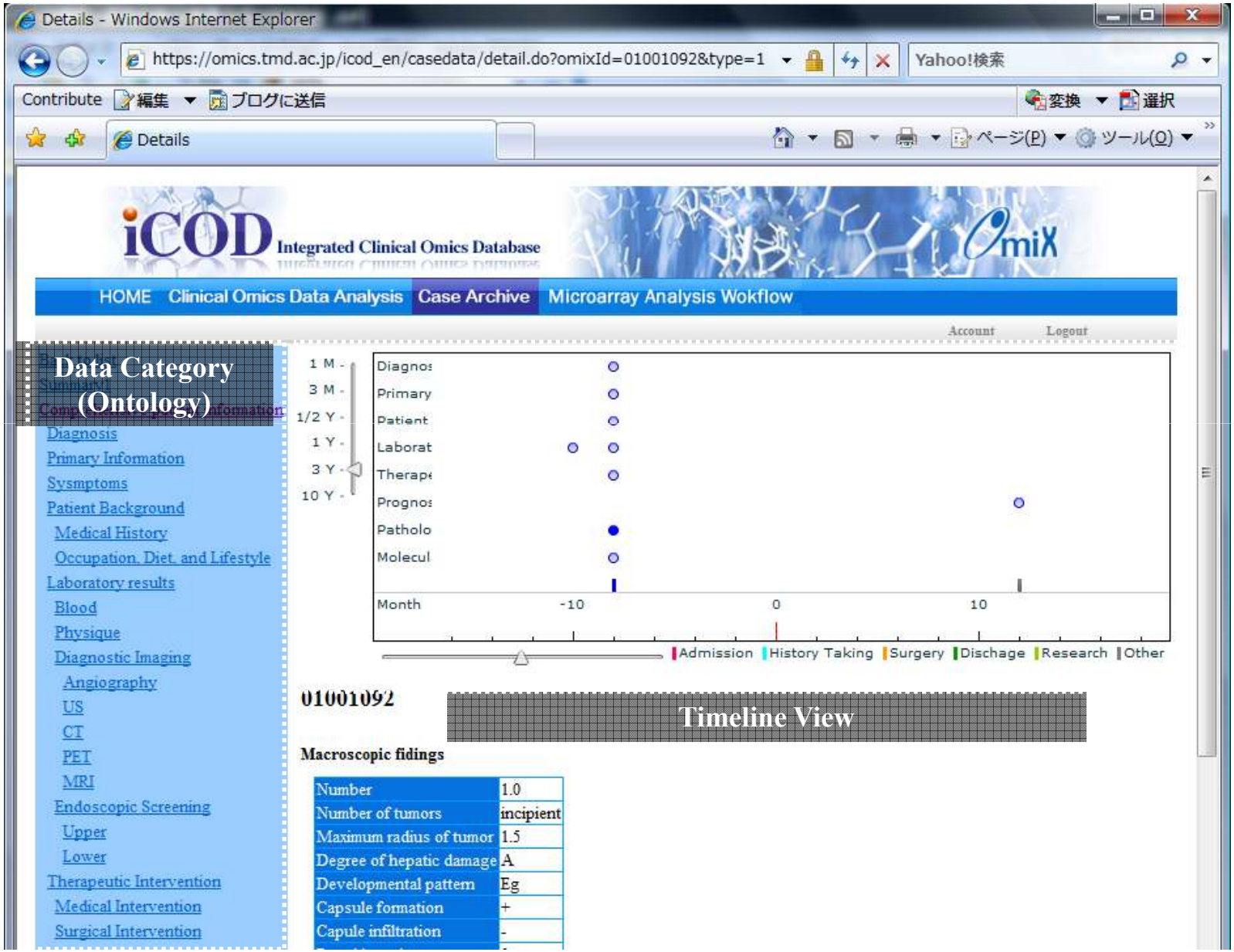
Top page

Integrated Clinical Omics Database - Windows Internet Explorer
https://omics.tmd.ac.jp/icod_en/portal/top.do
Integrated Clinical Omics Database
iCOD Integrated Clinical Omics Database
OmiX
HOME
Clinical Omics Data Analysis
Case Archive
Microarray Analysis Wokflow
Account
Logout
Clinical Omics Data Analysis
Case Archive
Microarray Analysis Workflow
Databsae for Translational Research
Center for Information Medicine, Tokyo Medical and Dental University has developped"integrated Clinical Omics Database (iCOD)" aiming to establish the basis of Omics-based Medicine and Systems Pathobiology.
We have lauched this project since 2005 with the support of Japan Science and Technology Agency and Ministry of Education, Culture, Sports, Science and Technology.
We have collected more than 500 cases of hepatic, colon, and oral cancer with 300 RNA expression analyses and 200 DNA copy number analysis with clinical information.
We opened Japanese version in July 2008 and English version will be availabel in April 2009.
About Project
Organization
Microarray Analysis Workflow
News
2009/3/23 English site will be available in April 2009
Privacy Policy
Terms of Use
Contact us
Center for Information Medicine
RIKEN
OmicSpace.RIKEN.jp
JBIRC
GeMDBJ
Hitachi GP
HitachiSoft Life Science
National University Corporation
Tokyo Medical and Dental University
https://omics.tmd.ac.jp/icod_en/map/top.do
ローカル イントラネット | 保護モード: 有効
100%

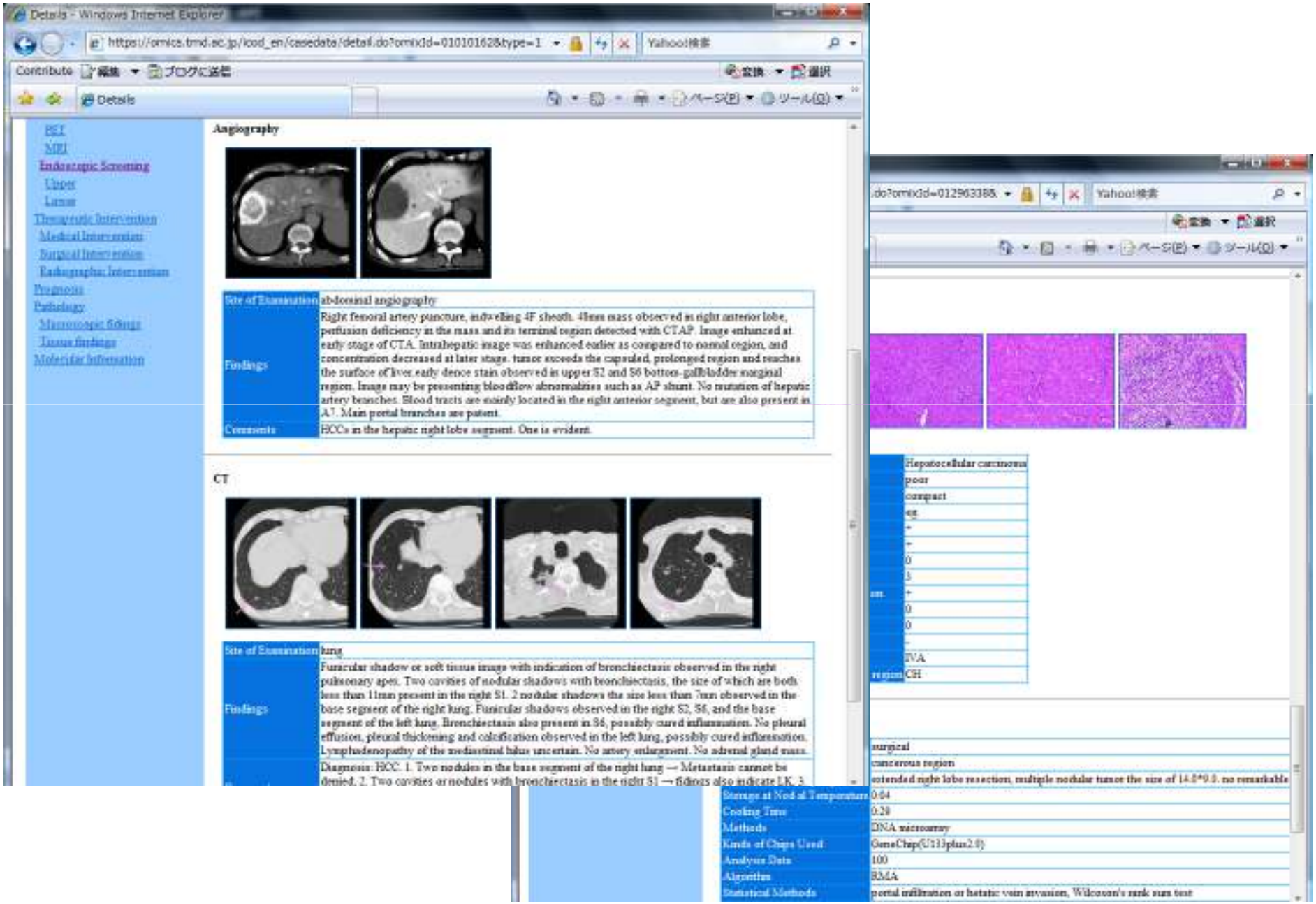
Case List

Case Archive - Windows Internet Explorer
https://omics.tmd.ac.jp/icod_en/casedata/reset.do?type=1&search=false
Yahoo!検索
Contribute
編集
ブログに送信
変換
選択
Case Archive
ページ(P)
ツール(O)
iCOD Integrated Clinical Omics Database
OmiX
HOME
Clinical Omics Data Analysis
Case Archive
Microarray Analysis Wokflow
Account
Logout
<<prev
Results 1 - 25 of 140
next>>
25
1: 01001092 Hepatocellular carcinoma
Treatment
heatic S5 subsectmental resection + S6 partial resection
Pathological Information
T=T1,N=N0,M=M0
2: 01010162 Hepatocellular carcinoma
Treatment
hepatic S8 subsegmental resection
Pathological Information
T=T3,N=N0,M=M0
3: 01017648 Hepatocellular carcinoma
Treatment
hepatic right lobe resection
Pathological Information
4: 01018257 Hepatocellular carcinoma
Treatment
hepatic S5+ 8 resection (right portal vein embolization)
Pathological Information
T=T3,N=N0,M=M0
5: 01036382 Hepatocellular carcinoma
Disease
All
Hepatocellular carcinoma
Colorectal cancer
Head and Neck Cancer
Esophageal cancer
Search:
Keyword:
Disease:
All
Category:
Item:
Value:
Condition:
Time:
NEW
AND
Search
Simple Search

Case Details (1)

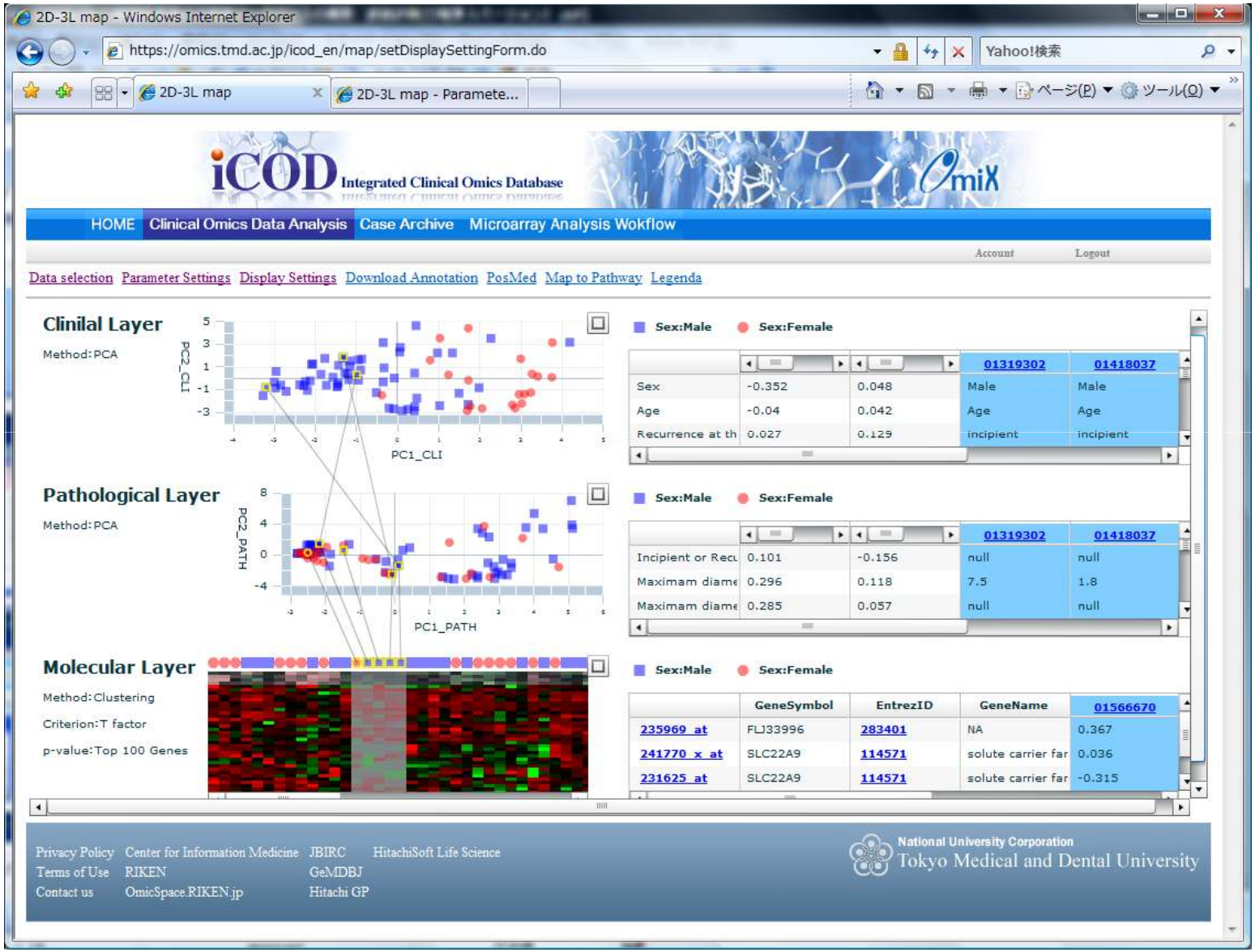


Case Details (2)



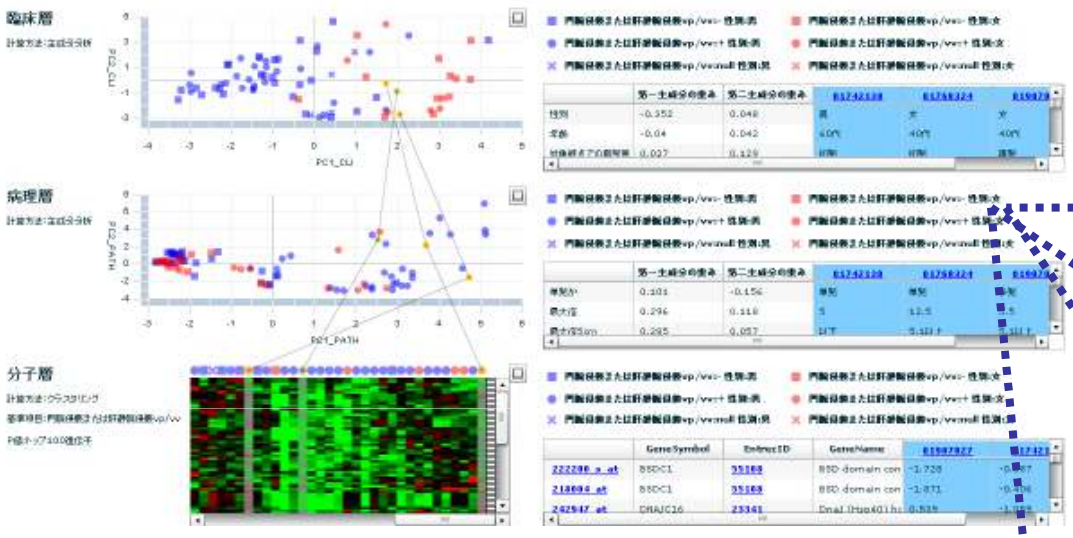
Clinical Omics Analysis

Clinical 3-layered Omics Map

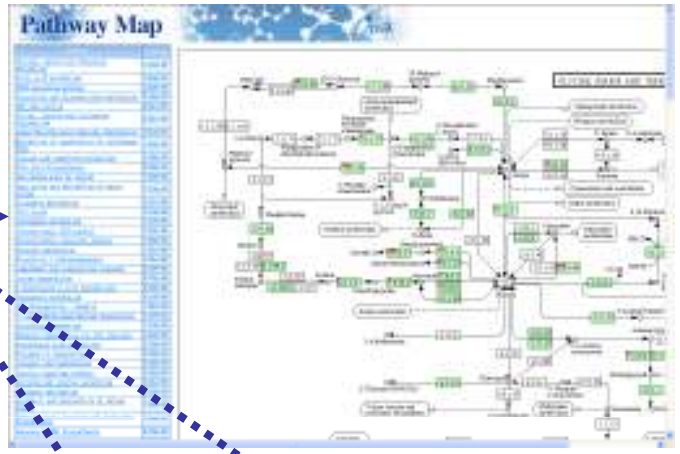


Omics data analysis

Three Layered Map

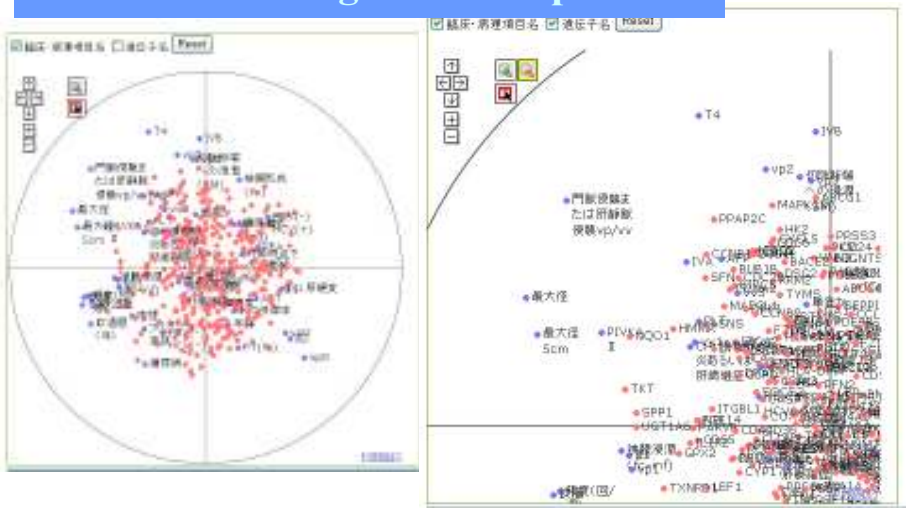


Pathway map



Transcriptome mapped on KEGG

Pathome-genome Map



Legenda

PosMed

OmicDownload

Semantic Navigation System

We show the semantic navigation system

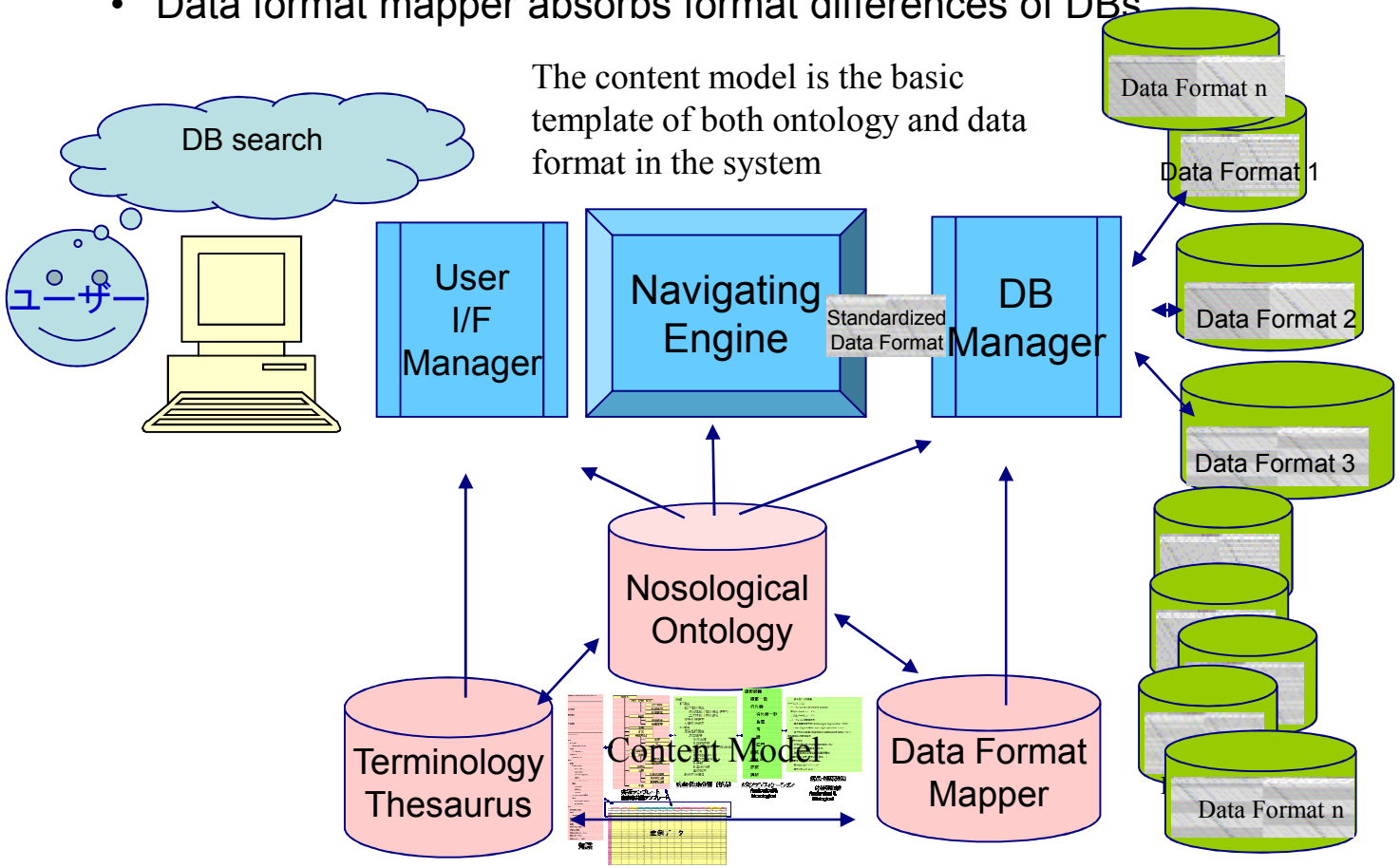
which we can say as a content interface based on the content model.

Integrated DB Guiding System

- for Core DB （**2 DBs＋α（10 DBs）** ： **Now in investigation**）

  We are developing **Searching System based on the ontology**

  - User I/F Manager convert user input to standardized word with terminology and thesaurus.
  - Navigating engine navigates users to targeted databases semantically.
  - Data format mapper absorbs format differences of DBs.



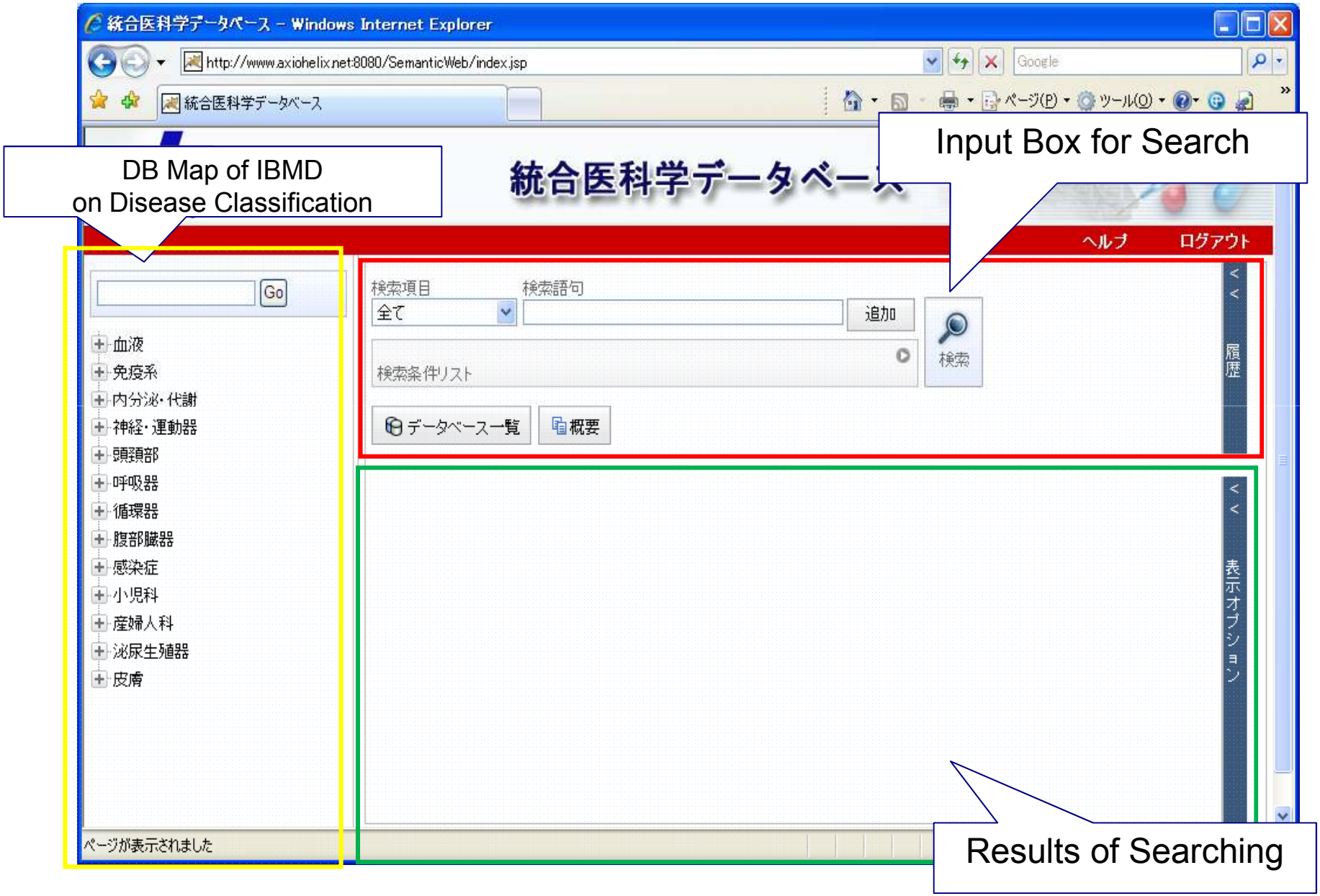
Prototype Demo

PPT version

IBMD Top Page

http://ibmd.tmd.ac.jp

Search Window



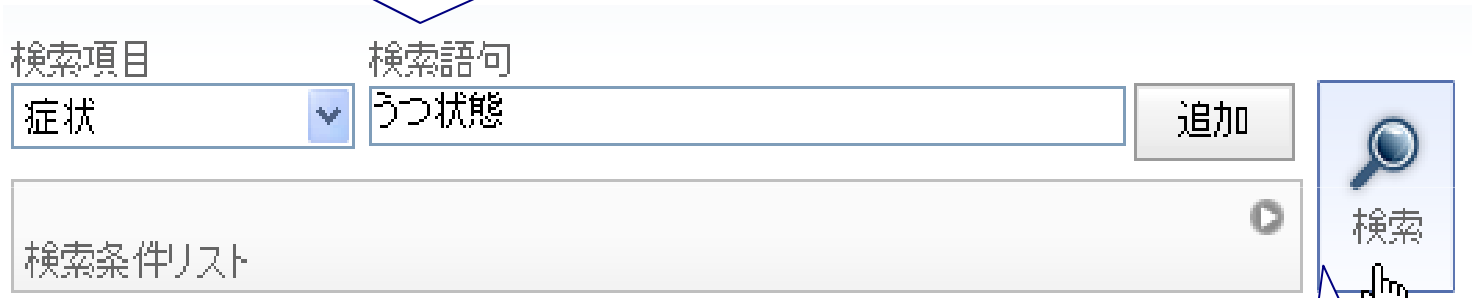
Example

– Inquiry “**Symptom：Depressive state**”

- Cancer DB in TMD
  - ＜D＞ 0 cases
  - ＜S＞ by Thesaurus, semantic transform
    “**Depressive state → Anorexia**” ⇒ **2 cases**
  - ＜S＞ “**Depressive state → Lassitude**” ⇒ **4 Cases**
    » one case is doubled 5 cases
- Parkinson Disease DB in Osaka Univ.
  - ＜D＞0 cases
  - ＜S＞Ontology、”**Depressive state → slow movement**”
  - ⇒ **6 cases**

Input

1．Category:”**Symtoms**”、Search Input Word:”**Depressive state**”



2．Push
Search Button

Results

データベース一覧　概要

1 - 11/11　First　Last

データベース名	病名	症状	S/D	翻訳	症例詳細
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
網羅的疾患分子病態DB	肝細胞癌	食欲不振	S	うつ状態 -> 食欲不振	
網羅的疾患分子病態DB	肝細胞癌	食欲不振、からだのだるさ	S	うつ状態 -> 食欲不振	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	

1. From **TMD DB**

<Direct> O cases

<Semantic>「Symptom: Depressive state → **Anorexia**」**2**cases

<Semantic>「Symptom: Depressive state → **Lassitude**」**4**cases

Duplicate case 1 case → （**Total 5 cases**）

2. From **Osaka Univ. DB**

<Direct> 0 cases

<Semantic> **Symptom: Depressive → Slow Movement** 6 cases

Detailed Data

Click here for detailed data

データベース一覧　概要

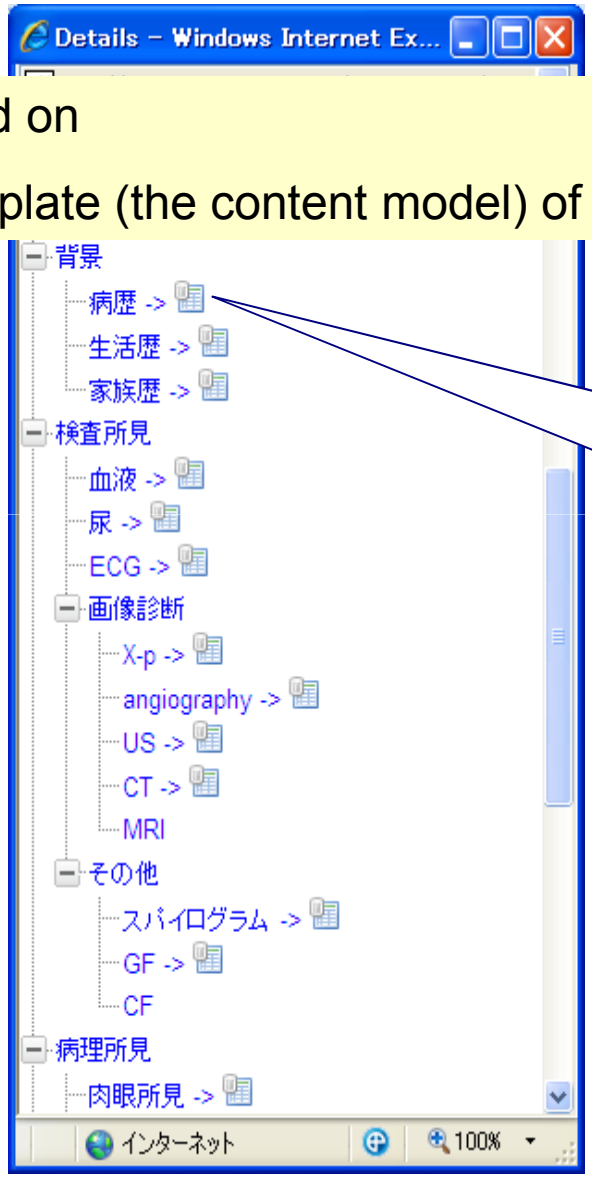
1 - 11/11　First　Last

データベース名	病名	症状	S/D	翻訳	症例詳細
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
パーキンソン病DB	パーキンソン病	動作緩慢	S	うつ状態 -> 動作緩慢	
網羅的疾患分子病態DB	肝細胞癌	食欲不振	S	うつ状態 -> 食欲不振	
網羅的疾患分子病態DB	肝細胞癌	食欲不振、からだのだるさ	S	うつ状態 -> 食欲不振	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	
網羅的疾患分子病態DB	肝細胞癌	からだのだるさ	S	うつ状態 -> からだのだるさ	

Detailed Data sub Window

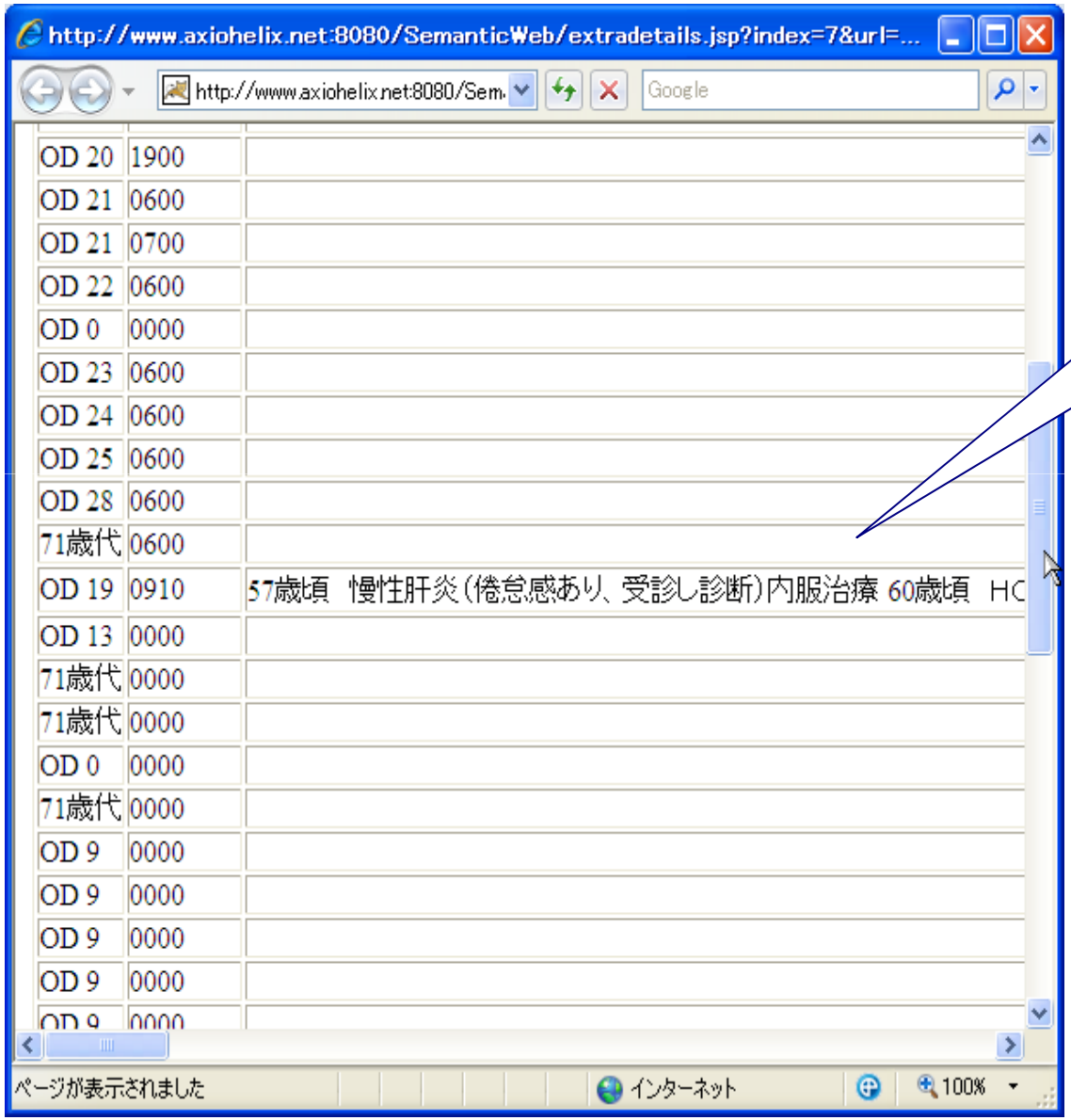
This panel is based on

the archetype template (the content model) of disease



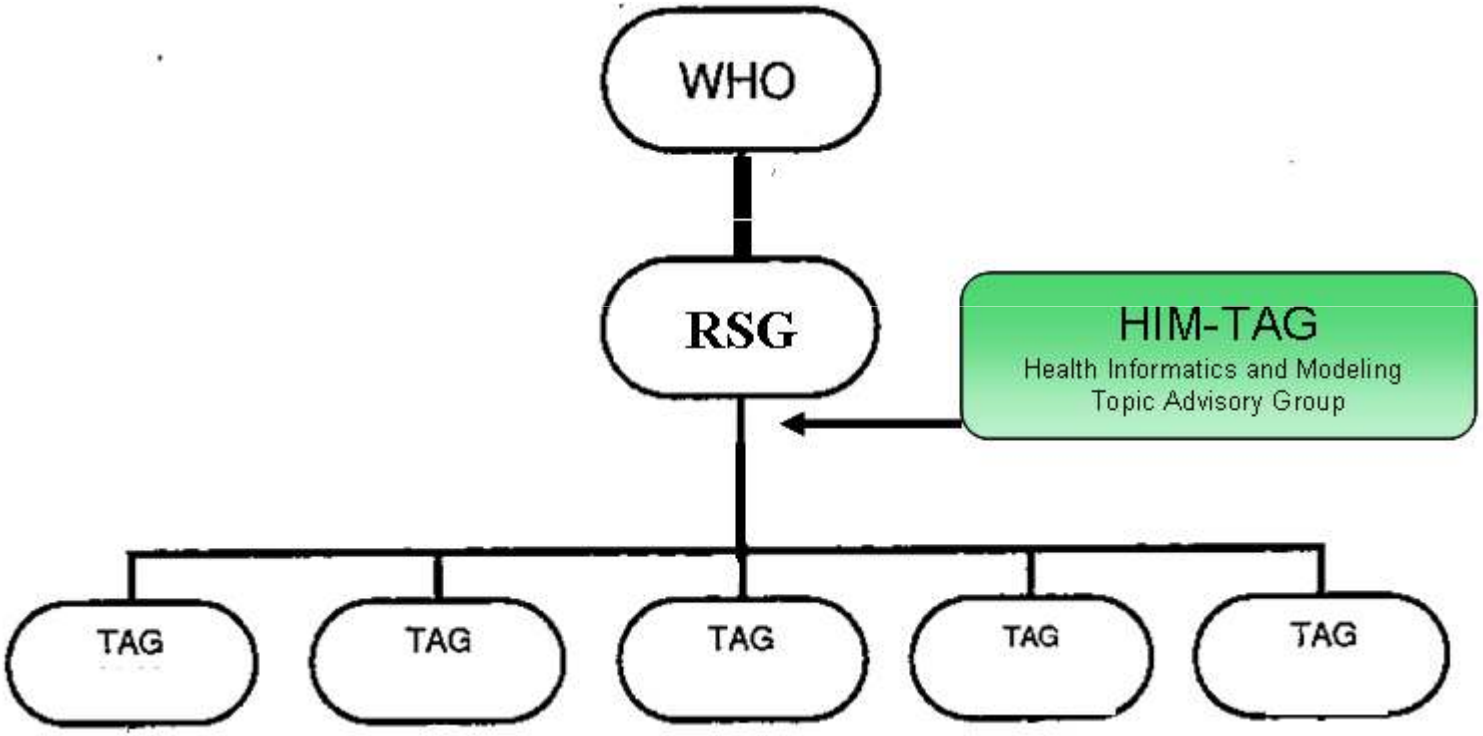
Hitted Patient Information is displayed according patient content model (template)

Patient Content



Ex：Patient History sub window

ICD11 and HIM-TAG



TAG-HIM present status

- 1. Content Model Group
  - Treat the original use case of ICD11 (Stefany)
- 2. Information group
  - Alan Rector (Univ. of Manchester, UK), John, Chris, SCT
- 3. Content Model front end for each TAG,
  - Categorial Structure for Rare disease
    - Jean Marie with Rare Disease TAG
  - IBMDB model is modified for internal medicine
    - Jun Nakaya and Hiroshi Tanaka with Internal Medicine TAG
  - Robert will take care of these things
- 4. SCT(SNOMED) Coordination
  - IHTSDO harmonization panel will take this issue.
    - (Kent, Alan Rector (Univ. of Manchester, UK), Chris, Olivier)
  - GO and other ontologies will cover the remained area.

Data Exchange Format
（as an Info. Model）

Genomic Sequence Variation Markup Language (GSVML)
ISO/#25720 Passed IS ballot

US, UK, Canada, Korea, Italy, Israel, Australia, Japan

Led by Jun Nakaya

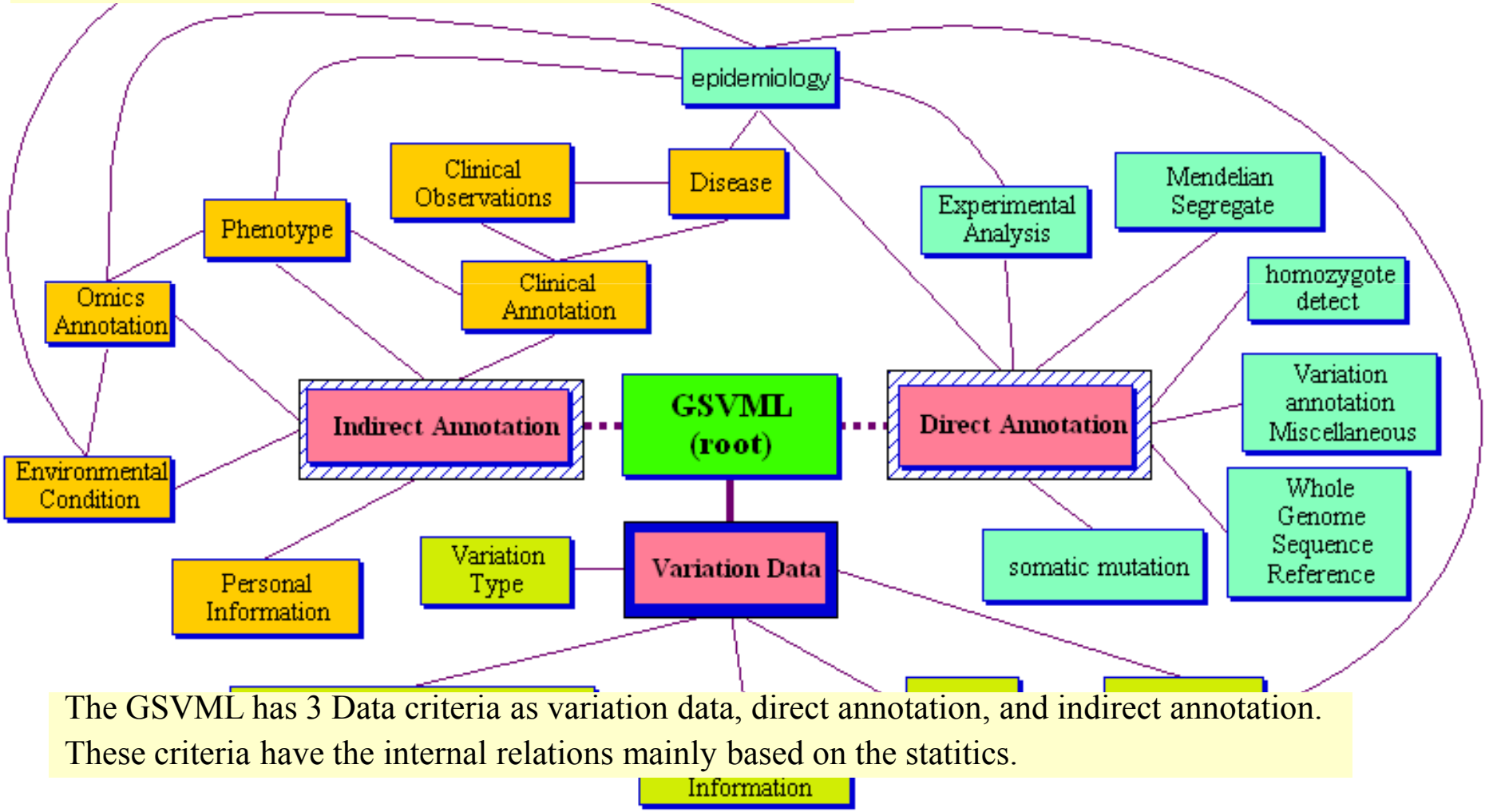
5th IBIC symposium Jun Nakaya, Michio Kimura, Hiroshi Tanaka

Nakaya, J., Hiroi, K., Yang, W., Ido, K., Kimura, M. (2006) "Genomic Sequence Variation Markup Language (GSVML) for Global Interoperability of Clinical Genomics Data(**Best Paper Award**)". Asia Pacific Association for Medical Informatics 2006 Proceedings. A01. 1-8.

Outlined Structure of GSVML

Nakaya, J., Hiroi, K., Yang, W., Ido, K., Kimura, M. (2006) (Best Paper Award) APAMI2006 A01: 1-8.

The GSVML has hierarchical structure.
The entry point of GSVML is genomic sequence variation.



The GSVML has 3 Data criteria as variation data, direct annotation, and indirect annotation.
These criteria have the internal relations mainly based on the statitics.

Thank you for your attention

Hiroshi Tanaka
(Director)
TMDU Biomedical Ontology Group
Jun Nakaya (Leader)
Keisuke Ido
Kaei Hiroi