

第25回 HL7 セミナー

臨床ゲノム情報学分野における社会情報基盤  
と  
GSVML

中谷 純

木村 通男

田中 博

東京医科歯科大学

浜松医科大学

# Domain Analysis on Clinical Genome Informatics (Clinical **Genomics**)

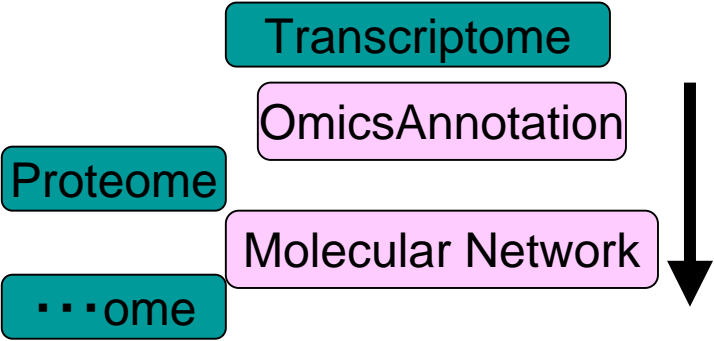
Nakaya, J., 2007, Clinical Genome Informatics (CGI) and its Social Information Infrastructure, IJCSNS, 7(1), 55-59

Shigetaka Asano, Jun Nakaya: Symposium on Genome Medical Informatics, Key Note, 2004

臨床ゲノム情報学  
= ゲノム情報に基づいて臨床医学(ゲノム医療)を支援するための情報学

ゲノム医療 = ゲノム情報に基づいて行う医療

- ゲノム (Gene + ome) = 多数の遺伝子の集まり
- ゲノム情報 (Genomics) = 遺伝子情報 + 相互関係情報



階層的オミックス情報  
= 各オミックス情報  
+ 相互関係情報 (階層内情報 + 階層間情報)  
+ ネットワーク総体情報 (総体としての振舞)

人体 (Phenome 情報) = 人での実質的発現系 正常



ゲノム医療

情報 = データ + 知識 (構造、Terminology, 分類)

2015 オミック医療一部現実化

2010 ゲノム医療一部現実化

法整備  
社会的認知機構  
統合データベース要件  
個人情報問題  
生物倫理問題  
DNAバンク機構  
遺伝子拡散および選択問題  
ゲノム情報機構  
知識発見理論の発展  
ゲノム情報の体系化

背景

数多くの  
解決すべき課題

2004 FDA Innovation or Stagnation

2003 NIH Road Map

ポストゲノム時代

2003 ヒトゲノム配列解析完了宣言

# ゲノム医療の 三つの特徴

「**個**」の医療

⇒ 国ごと、地域ごと、個人ごとの医療

「**先手**」の医療

⇒ 「予測・予防」の医療

「**科学的**」医療

⇒ 分子的証拠に基づく医療

創薬、個別化医療、経済的医療

現実世界への適用が主眼

# 将来医療のゴール(方向性)

SGMI 2004 S. ASano, J. Nakaya, H. Tanaka

ゲノム医療が主役になると思われる将来医療

- 個人差を考慮した医療
- 疾病予防

共通点 = 予測に基づく  
→ 多量の正確な情報(ゲノム・オミックス)  
と予測技術 が必要

ゲノム医療が脇役になると思われる将来医療

- 無痛医療
- 心まで目を向けた全人医療

Copyright to S. ASano, J. Nakaya

All Rights to Jun Nakaya

# ゲノム医療における情報学の役割

=3つの臨床目的を達成するための情報学的支援

3つの臨床目的 (優先順位順)

- 1 "臨床安全性" 向上
- 2 "臨床効果" "臨床作業効率" 向上
- 3 "臨床経済性" 向上

医療の分野で、これまで情報学は主に経済性(3)を扱ってきた

多量の情報(オミックス、ゲノム) ↓ を扱う時代となり

臨床医療そのもの(1,2)を科学する情報学が可能となった

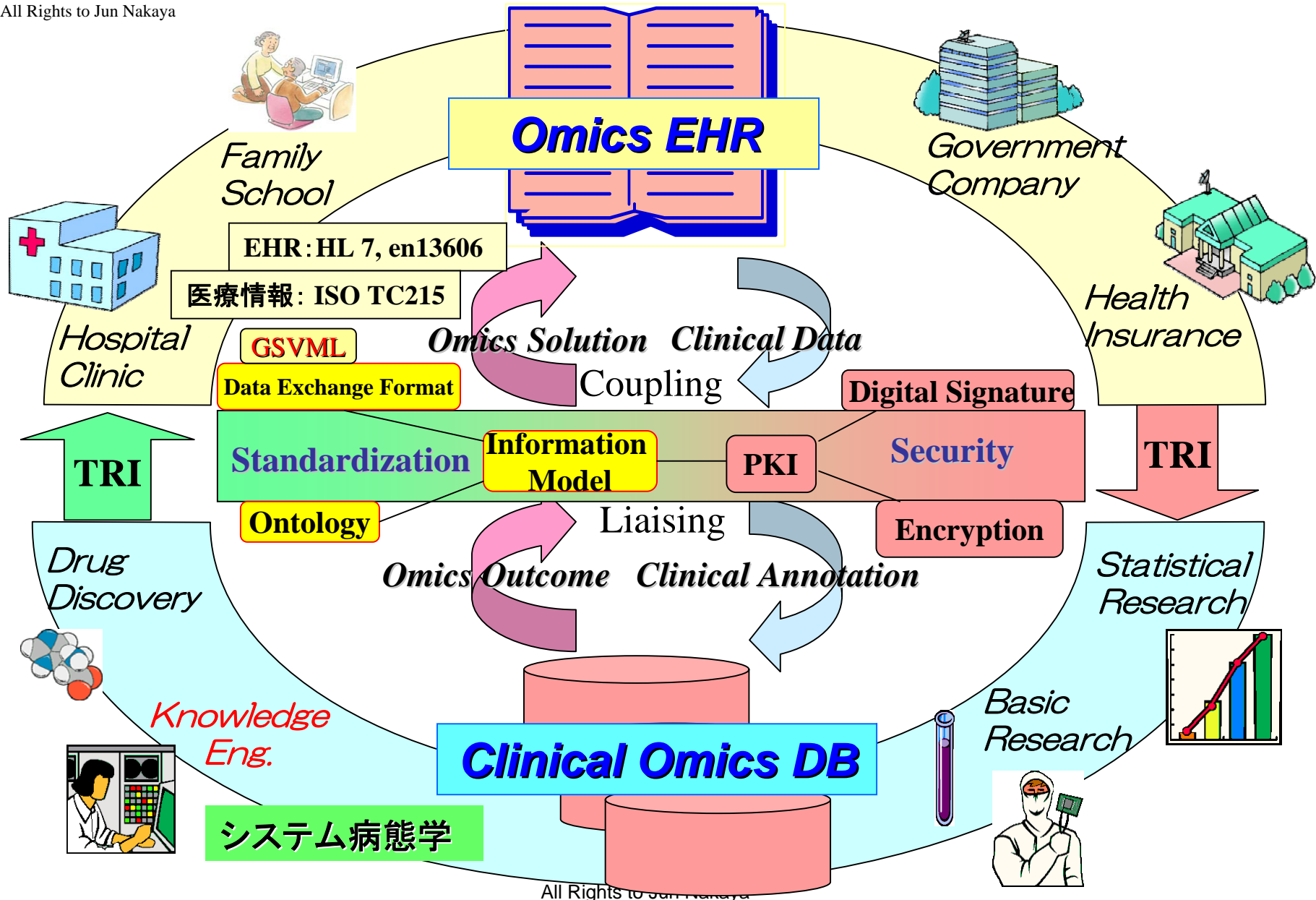
# ゲノム医療に必要な 基盤情報インフラ

平成18年度 産総研 TRI研究会 基調講演、「これからの医療と病院のあり方 - TRI、CGI、COIの視点から -」、中谷 純

- 社会情報基盤
  - 標準化（情報の共有、交換、有効利用）
  - セキュリティ（個人情報保護）
  - トランスレーショナルリサーチ情報基盤
    - ゲノム医療現実化への社会的パイプライン
      - （基礎研究成果を社会へつなげるための情報基盤）

# ゲノム医療の社会情報基盤

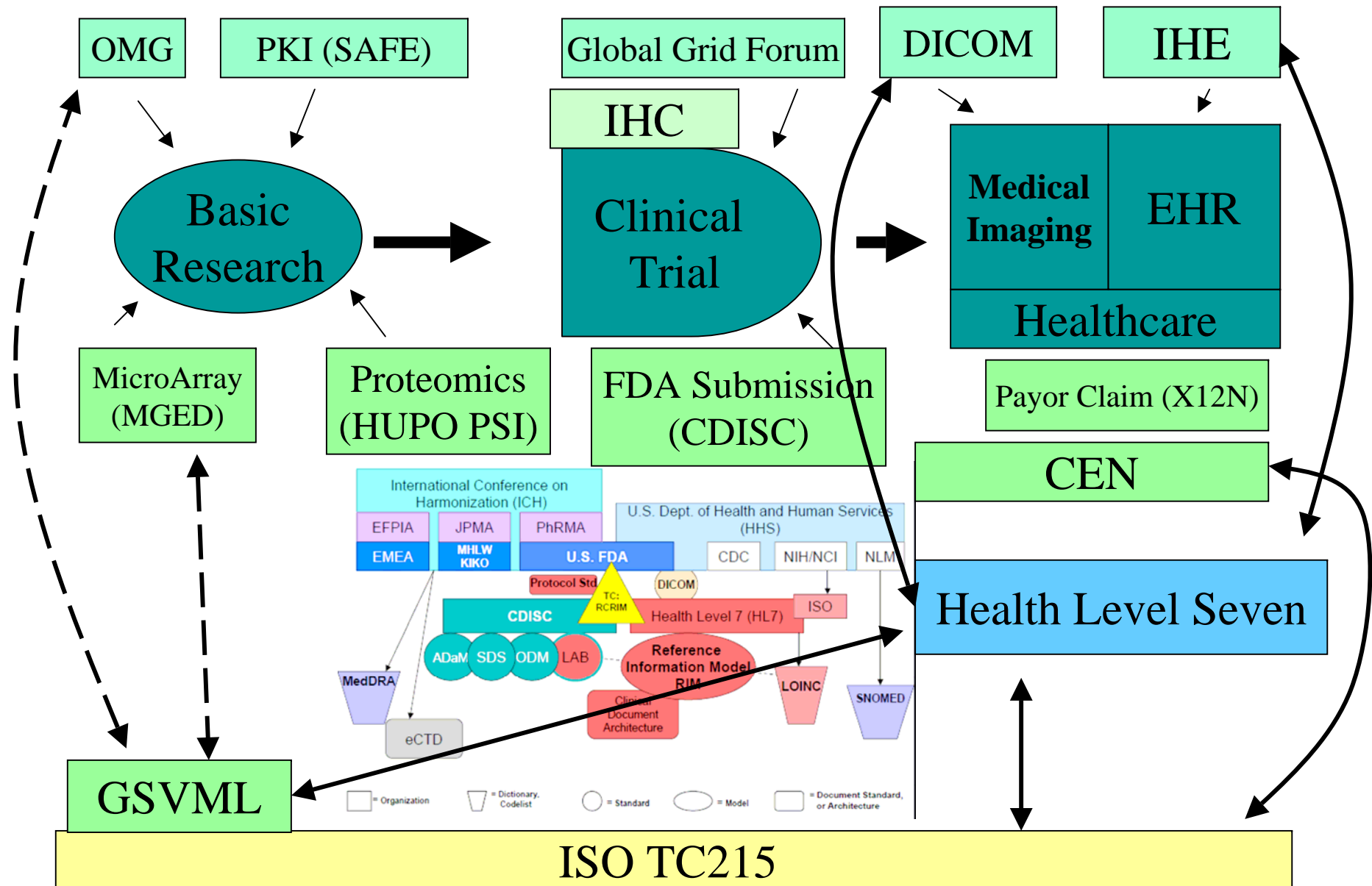
All Rights to Jun Nakaya



All Rights to Jun Nakaya

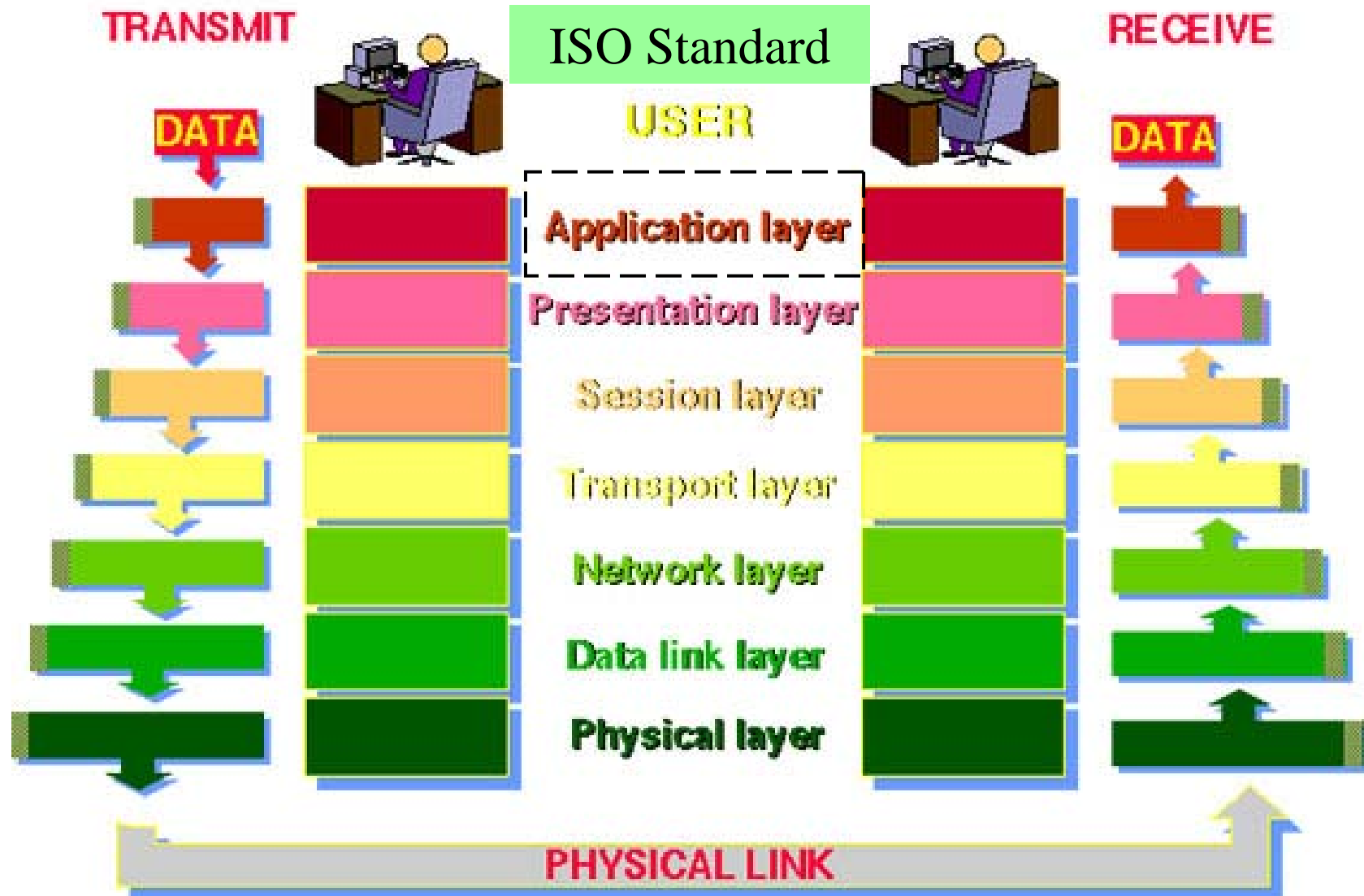


# Healthcare and Life Science Standardization Bodies



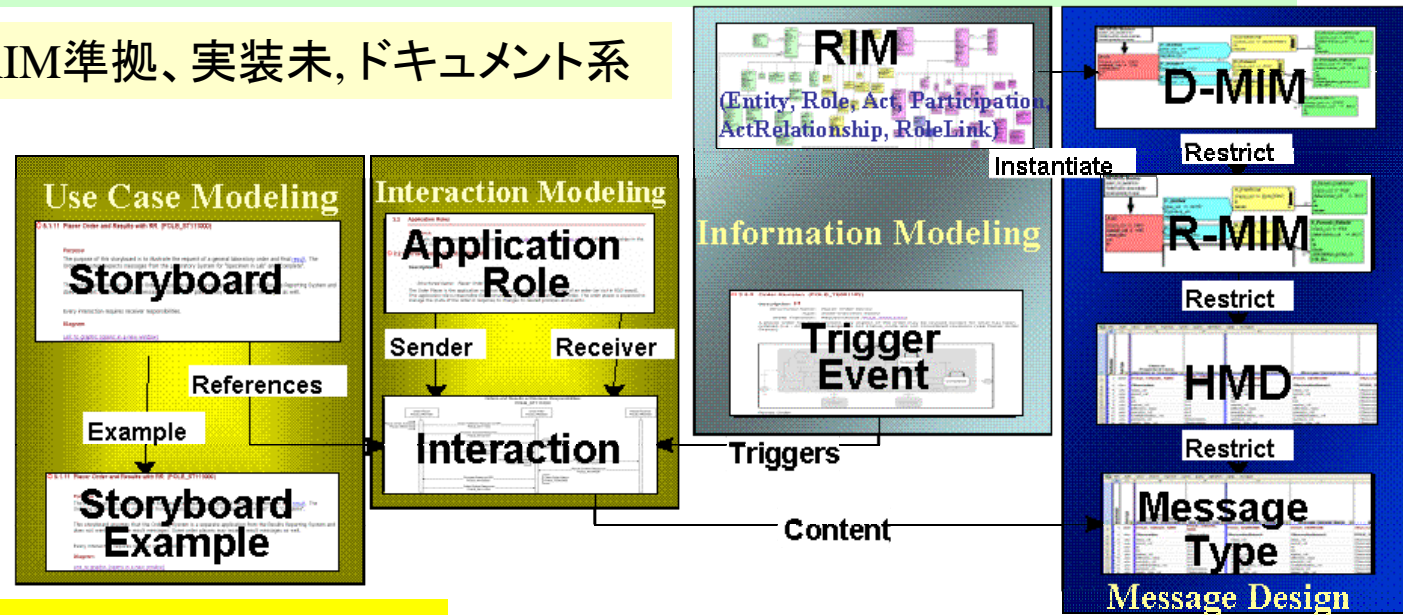
# HL7 is THE 7 LAYERS OF OSI

Open System  
Interconnection



# HL7の情報モデル

- v3 : XML形式RIM準拠、実装未、ドキュメント系

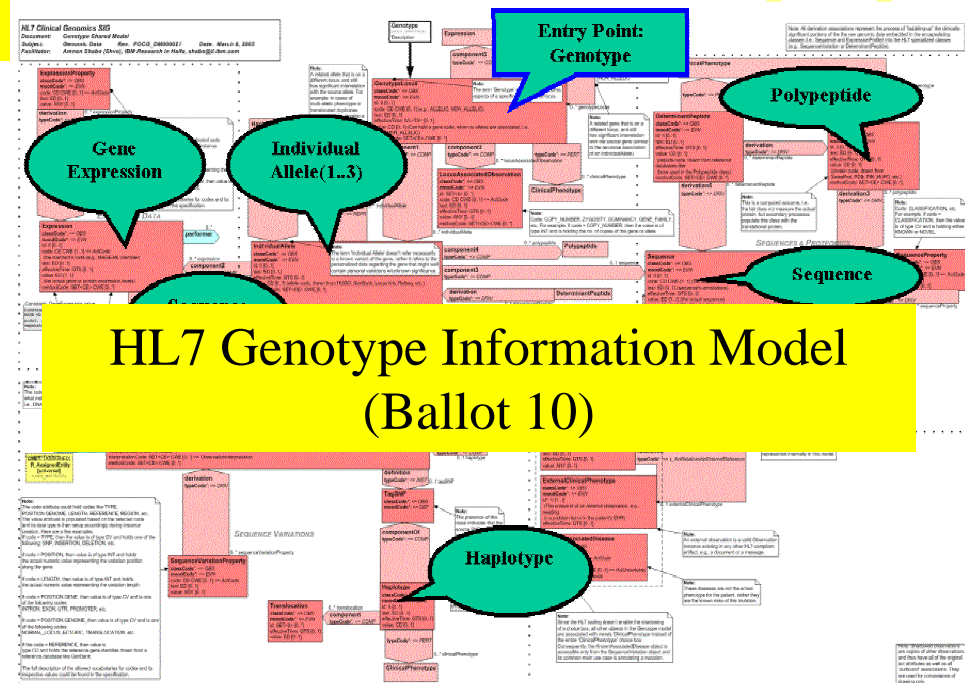


オントロジーは、ボキャブラリーとして利用

v2.5とv3は、今後も並存の予定

- v2.5: Encoding Rule 7形式、実装済、トランザクション系

- 厚生労働省標準的電子カルテ



## HL7 Genotype Information Model (Ballot 10)

```
MSH... <cr>
PID|1||123456789||Levin^Henry^^the
7th||19230513<cr>
OBR|1||^Chest X-ray<cr>
```

copyright to HL7 CG SIG

All Rights to Jun Nakaya

# ISO Outline

TC 215/CAG 1 Executive council, harmonization and operations

The convener can be reached through:

TC 215/WG 1 Data structure

The convener can be reached through: SCC

TC 215/WG 2 Data interchange

The convener can be reached through: ANSI

**TC 215/WG 3 Semantic content**

**The convener can be reached through: ANSI**

TC 215/WG 4 Security

The convener can be reached through: SCC

TC 215/WG 5 Health cards

The convener can be reached through: DIN

TC 215/WG 6 Pharmacy and medicines business

The convener can be reached through: NEN

TC 215/WG 7 Devices

The convener can be reached through:

TC 215/WG 8 Business requirements for Electronic Health Records

The convener can be reached through: SA

ISO is

- a network of the national standards institutes of 156 countries
- a non-governmental organization
- many of its member institutes are part of the governmental structure of their countries

**TC 215**

**Health informatics**

Secretariat: ANSI

Secretary: Ms. Audrey Dickerson

Chair: Dr. Yun Sik Kwak (Korea) until end 2009

Scope: Standardization in the field of information for health, and Health Information and Communications Technology (ICT) to achieve compatibility and interoperability between independent systems. Also, to ensure compatibility of data for comparative statistical purposes (e.g. classifications), and to reduce duplication of effort and redundancies.

# TR and Critical Path

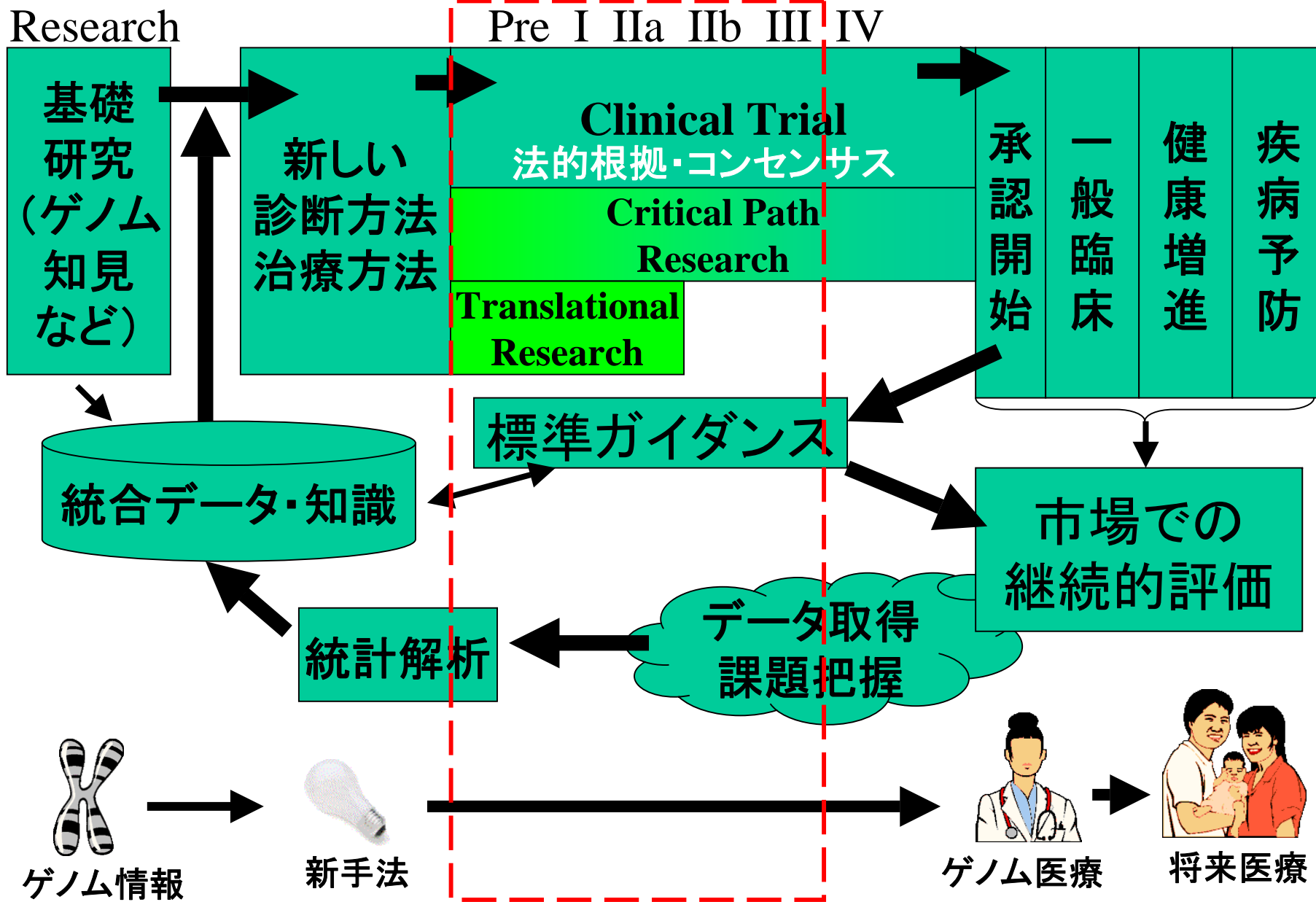
## トランスレーショナルリサーチ情報基盤

ー ゲノム医療現実化への社会的パイプライン

(基礎研究成果を社会へつなげるための情報基盤)

Nakaya, J. 2006 The Translational Research Informatics (TRI) (Leadoff Article), International Journal of Computer Science and Network Security, 6(7A), 117-122

# ゲノム医療とパイプラインの流れ

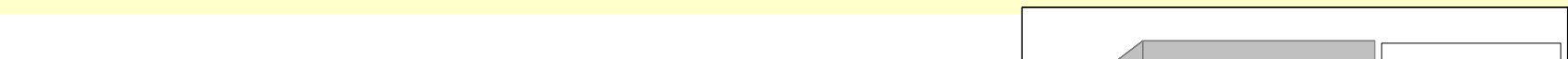


# クリティカルパス最適化に必要な情報基盤

NIH: Elias Zerhouni: The NIH Roadmap Science. 2003 Oct, Vol. 302: 3: 63-72

FDA: Innovation or Stagnation ? March 2004

- 医生物学研究コスト ↑
- FDAへの新規薬剤登録申請数 ↓
- 一つの薬剤あたりの申請までの開発費用 ↑  
1.1B \$ to 1.7B \$ (Critical Path に集中)



クリティカルパス  
(特に、Translational Research Phase)  
の最適化が必要

# 三つの基軸と重要技術概要

Phase	Basic Research	Prototype Design or Discovery	Preclinical	Clinical Trial	FDA Filing/Approval & Launch Preparation
基軸	安全性	材料選択 構造 活性 相関	試験管テスト 動物テスト	人体テスト 動物テスト	安全性フォローアップ
	医療有用性	実験確認 コンピューターモデル 確認	試験管モデル 動物モデル	人体有効性評価	有用性フォローアップ
	産業性	物理設計	特徴化 少量生産	特徴の改良 スケールアップ	特徴データの集積 大量生産

**重要**

- できるだけ早期における臨床結果の予測、計測技術



# 目的別必要例：血液疾患の場合 (縦軸：臨床フェーズ、横軸：病因)

RR2002

Phase	Example Disease	von Willebrand Disease	Acute Myeloblastic Leukemia	Adult T cell Leukemia/Lymphoma	Depression	
Basic	典型的症状 必要なもの					
	分子構造-機能連関表	vWF VIII因子関連遺伝子ネットワーク	Blast系細胞増殖と細胞死についてのネットワーク	骨髄球増殖性(線維化)についてのネットワーク	細胞性免疫(特に、Helper T Cell)と免疫反応についてのネットワーク	
Prototype Design or Discovery	構造-ヒト有効性関連データベース 物質活性-ヒト有効性についての知識ベース	vWF VIII因子そのものおよび産生因子とヒトでの有効性に関する知識ベース	増殖因子、抑制因子、細胞死因子とヒトでの影響結果についての事例	線維芽細胞増殖因子、コラーゲン沈着因子、髄外造血誘導因子とヒトでの影響結果についての事例	Helper T Cellを中心とした、免疫物質、免疫物質、カスケード変化とヒトでの影響結果についての事例	
	標的物質の毒性知識ベース	vWF前での特異蛋白質の特定とヒトでの毒性結果についての事例	増殖因子、抑制因子、細胞死因子に対するシグナルの特定とヒトでの毒性影響結果についての事例	巨核球成長因子、線維芽細胞シグナル、コラーゲンシグナルとヒトでの毒性影響結果についての事例	感染Helper T Cell標識、あるいは、Cell to Cell感染因子とヒトでの毒性影響結果についての事例	
Prototype Design & Preclinical	バイオシステム-臨床効果相関表	vWFトランスジェニックマウス、セルライン、セルトとヒトvWF患者との相関表	AMLトランスジェニックマウス、セルラインとヒトAML患者との相関表	ATLLトランスジェニックマウス、セルラインとヒトとの相関表	うつトランスジェニックマウス、セルラインとヒトとの相関表	
	モデルに基づく開発行程設計システム	vWF薬剤分子モデルデータベース統合 構造-機能(臨床効果) 連関経験的知識ベース構築	AML薬剤分子モデルデータベース統合 構造-機能(臨床効果) 連関経験的知識ベース構築	CIMF薬剤分子モデルデータベース統合 CIMFモデル	ATLL薬剤分子モデルデータベース統合 ATLLモデル	抗うつ薬 分子モデルデータベース統合 うつモデル
PreClinical	包括的開発結果予測システム 試験管内、動物、ヒトの間の因果関係に関する関連性の表	vWFにおける(試験管内-動物-ヒト)という因果関係についての知識ベース	AMLにおける(試験管内-動物-ヒト)という因果関係についての知識ベース	CIMFにおける(試験管内-動物-ヒト)という因果関係についての知識ベース	ATLLにおける(試験管内-動物-ヒト)という因果関係についての知識ベース	うつ状態における(試験管内-動物-ヒト)という因果関係についての知識ベース
	因果関係表を作るためのデータ収集機構	vWFについて、過去の実験データの掘り起こし作業(データベースの結合、データフォーマットの共通化)など	AMLについて、過去の実験データの掘り起こし作業(データベースの結合、データフォーマットの共通化)など	CIMFについて、過去の実験データの掘り起こし作業(データベースの結合、データフォーマットの共通化)など	ATLLについて、過去の実験データの掘り起こし作業(データベースの結合、データフォーマットの共通化)など	うつ状態について、過去の実験データの掘り起こし作業(データベースの結合、データフォーマットの共通化)など
PreClinical	人、動物 実験結果の間の因果関係についての知識ベース	vWFそのものおよび関連因子について、実験室、動物、ヒトの間の共通性およびその度合い(確率)についての因果関係知識ベース(事例収集データベース)	AMLそのものおよび関連因子について、実験室、動物、ヒトの間の共通性およびその度合い(確率)についての因果関係知識ベース(事例収集データベース)	ATLLそのものおよび関連因子について、実験室、動物、ヒトの間の共通性およびその度合い(確率)についての因果関係知識ベース(事例収集データベース)	うつ状態関連因子について、実験室、動物、ヒトの間の共通性およびその度合い(確率)についての因果関係知識ベース(事例収集データベース)	
	動物-ヒト有効性関連データベース セルライン-ヒト有効性関連データベース	ヒトとの関連性の明確化 関連の明確な動物モデル作成	vWF病-トランスジェニックマウスモデル AML動物モデル AML動物-ヒト関連知識ベース	免疫マウス、炎症マウス ヒトとの関連知識ベース	精神疾患マウスモデル マウス-ヒト精神の関連知識ベース	
PreClinical & Clinical	簡易遺伝子導入法 生細胞イメージング(ハードウェア(KSR, 光, fMRI, PET)	セルラインの目的別、疾病別 簡易ツールキット化 実験室(生細胞-セルライン) 人体内	KSR, 光-マウス-ハードウェア fMRI, PET-マウス-ハードウェア	TR標準プロトコル AML TR標準プロトコル 標準プロトコルの個別化ルール	精神疾患TR標準プロトコル 標準プロトコルの個別化ルール	
	プロトコル作成システム プロトコル実行管理システム プロトコルデータベース記録と知識ベース化	vWF TR標準プロトコル 標準プロトコルの個別化ルール プロトコル実行力適応 プロトコルの作成管理 プロトコルの実行管理(適応)	AML TR標準プロトコル 標準プロトコルの個別化ルール プロトコル実行力適応 プロトコルの作成管理 プロトコルの実行管理(適応)	免疫マウス標準プロトコル 標準プロトコルの個別化ルール プロトコル実行力適応 プロトコルの作成管理 プロトコルの実行管理(適応)	精神疾患標準プロトコル 標準プロトコルの個別化ルール プロトコル実行力適応 プロトコルの作成管理 プロトコルの実行管理(適応)	
Clinical	バイオシステム-臨床効果バイオマーカー	vWFにおける、バイオシステム(細胞、セルライン)の共通のバリエーションの作成	AMLにおける、バイオシステム(細胞、セルライン)の共通のバリエーションの作成	CIMFにおける、バイオシステム(細胞、セルライン)の共通のバリエーションの作成	うつ状態における、バイオシステム(細胞、セルライン)の共通のバリエーションの作成	
	安定的かつ連続的な免疫状況計測手法	相関 相関に対する 連続的かつ簡便 信頼できる手法(高感)	AMLでの免疫物質 連続的かつ簡便 信頼できる手法(高感)	免疫マウスでの免疫物質 連続的かつ簡便 信頼できる手法(高感)	うつ状態における、免疫物質 連続的かつ簡便 信頼できる手法(高感)	
Launch	モジュール型生産システム 標的物質別バイオマーカー表 臨床反応別バイオマーカー表	vWF 病団体形成 TR対象疾患のカテゴリズと患者さんのとりまとの 生産システムのコンパチビリティと最適組み合わせを可能とするシステム vWF因子そのものでなく、それがもたらす臨床反応(出血)の物質マーカー データフォーマットが統一 共通のプラットフォーム仕様 データ収集協力コミュニティー 柔軟な設計変更が可能な生産工程システム	AML 病団体形成 TR対象疾患のカテゴリズと患者さんのとりまとの 生産システムのコンパチビリティと最適組み合わせを可能とするシステム 腫瘍マーカーそのものでなく、薬を投与した際の臨床反応と作用、副作用の因果関係を示すマーカー	CIMF 病団体形成 TR対象疾患のカテゴリズと患者さんのとりまとの 腫瘍、門脈圧亢進などの各項目に対するバイオマーカー	うつ病団体形成 TR対象疾患のカテゴリズと患者さんのとりまとの うつ状態に対応したバイオマーカー、あるいはイメージング	
	安全性データ、有効性データ、薬物性データのフォローアップ体制がある データ収集が容易なツール モジュール型生産設計手法					

PreClinical & Clinical	プロトコル作成システム	Neoplastic	Prototype Design or Discovery	モデルに基づく開発行程設計システム	Congenital
		Acute Myeloblastic Leukemia			
		<ul style="list-style-type: none"> <li>vWF薬剤分子モデルデータベース統合</li> <li>vWFモデル</li> <li>構造-機能(臨床効果) 連関経験的知識ベース構築</li> </ul>			

2004 Nov. S. Asano, J. Nakaya

# 情報学分野(CGI, TRI)への要請

1. 体系化: 情報収集と知識整理

標準化



- 用語、分類、知識指標として(Ontology)
- 標準データ交換フォーマット(GSVML)
- 統合データベース構築
- BIとCIの融合(CGI、システムパソロジー)

2. 予測手法の確立

- 新規治療法、新規診断法による効果、副作用、臨床リスク
- クリニカルパス(= 臨床的に対象患者様がたどる経路)
- 分子標的
- 計測技術(予測データの収集のため)

システム病態学



3. 予測精度評価のためのベンチマークの開発

4. 定量的予測のための医生物学知識定量化手法の確立

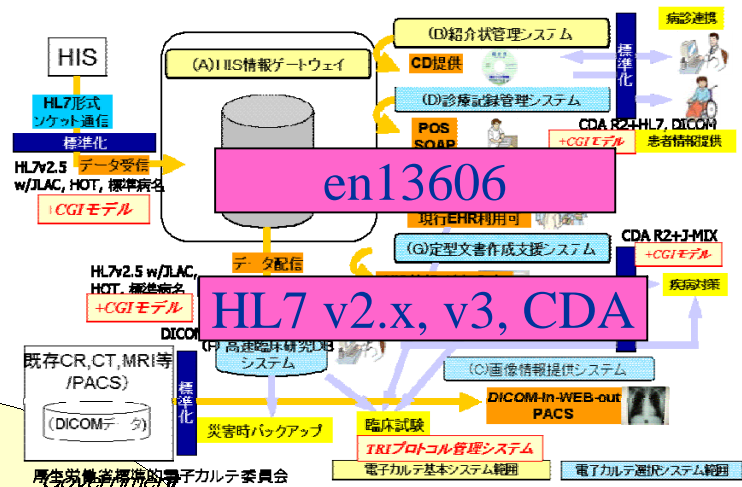
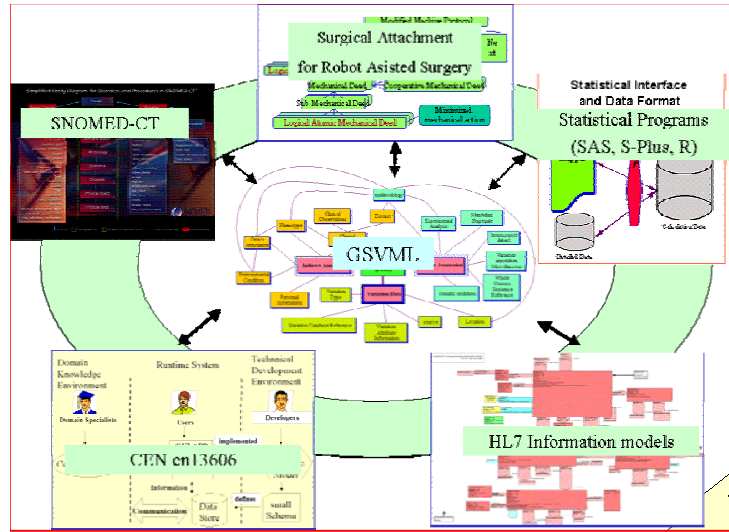
5. 個別医療のための、個のゲノム統計学

6. クリティカルパスにおける情報学的に最適化された臨床計画管理手法の開発

いわゆるプロトコル管理システム

All Rights to Jun Nakaya

# データフォーマット

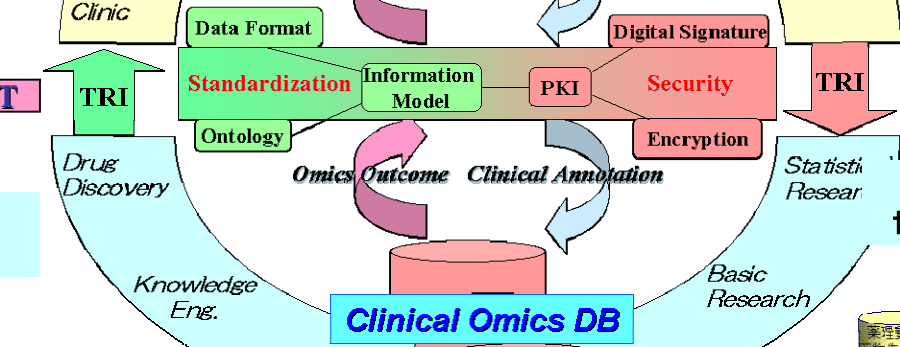


**国際標準化**  
世界初の臨床ゲノムデータ構造  
国際標準化プロジェクト

**オントロジー**  
臨床トップオントロジー  
SNOMED-CT

**システム病態学**  
病態のシステムの理解による予測技術

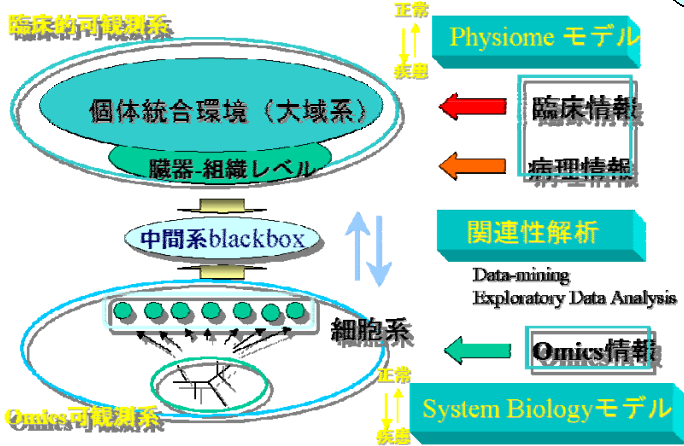
## ゲノム医療の社会情報基盤と関連プロジェクト



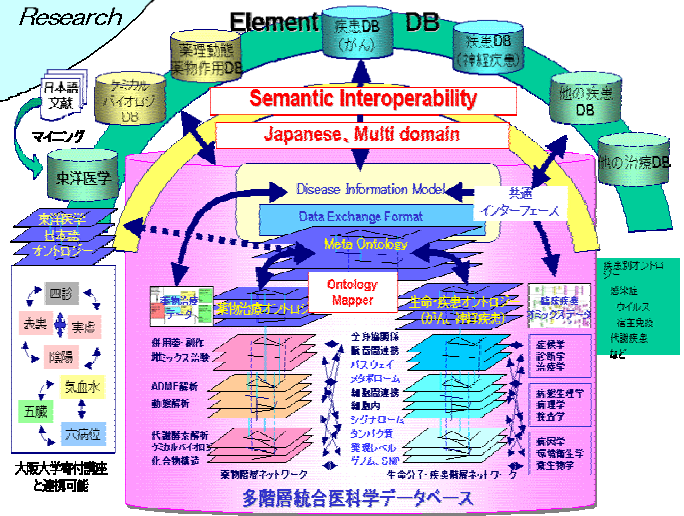
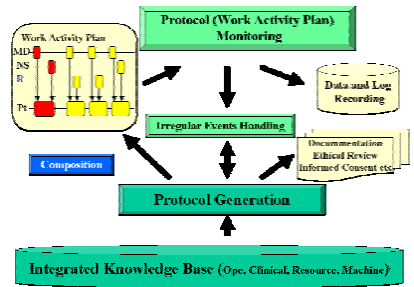
**生涯EHR**  
厚生労働省標準的国際規格のEHR

**情報モデル**

**統合データベース**  
世界最高水準の医生物学データベース



## プロトコル管理



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

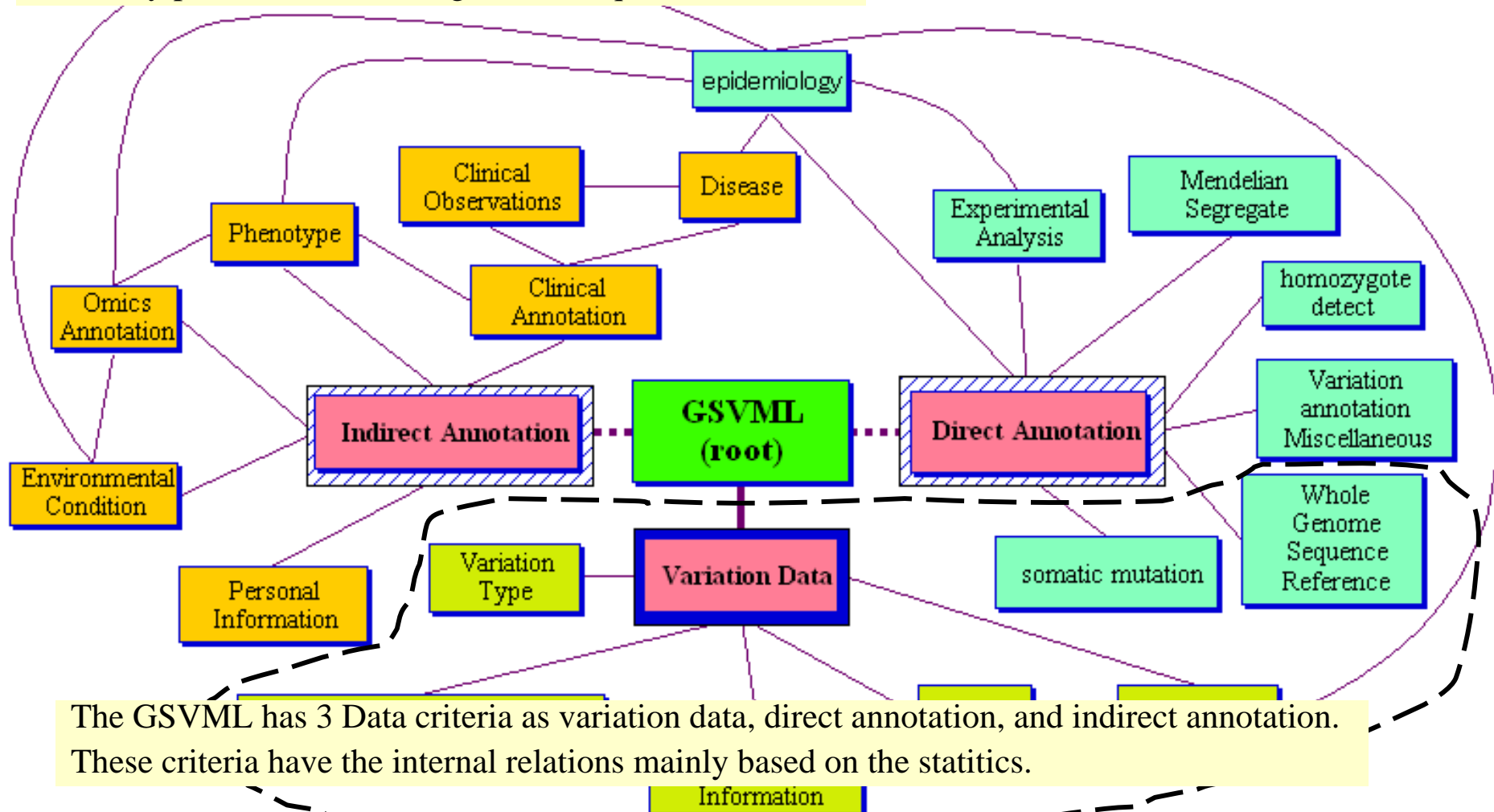
**US, UK, Canada, Korea, Italy, Israel, Australia, Japan  
Led by Jun Nakaya**

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

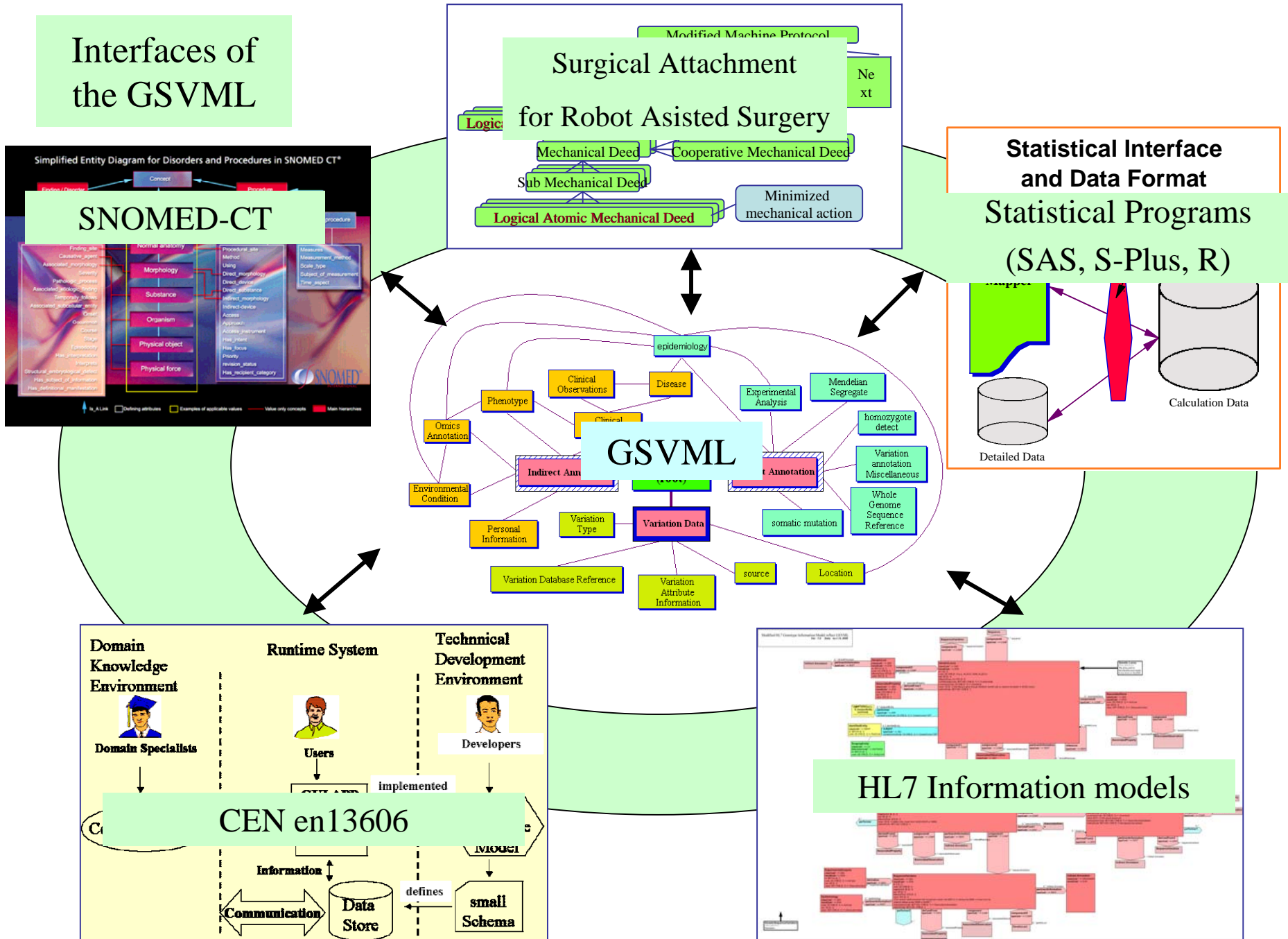
Nakaya, J., Hiroi, K., Yang, W., Ido, K., Kimura, M. (2006) (Best Paper Award) APAMI2006 A01: 1-8.

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

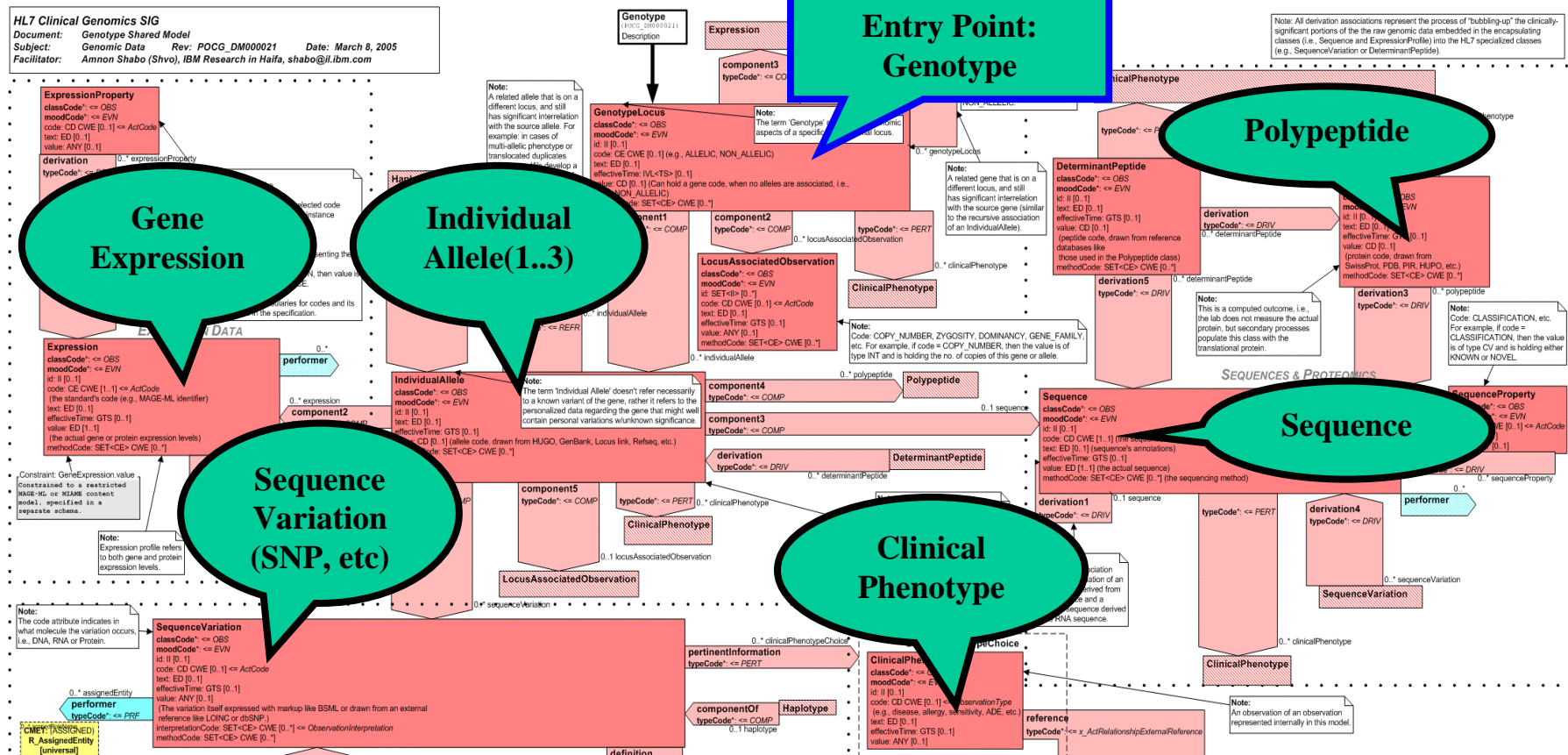


The GSVML has 3 Data criteria as variation data, direct annotation, and indirect annotation.  
These criteria have the internal relations mainly based on the statistics.

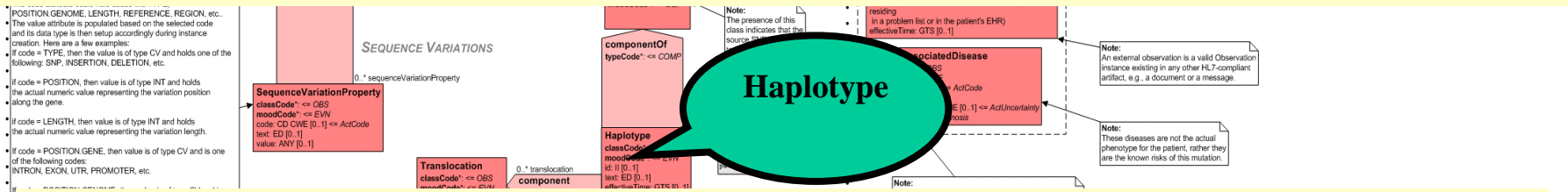
# Interfaces of the GSVML



# HL7 Genotype Information Model (Ballot 10)



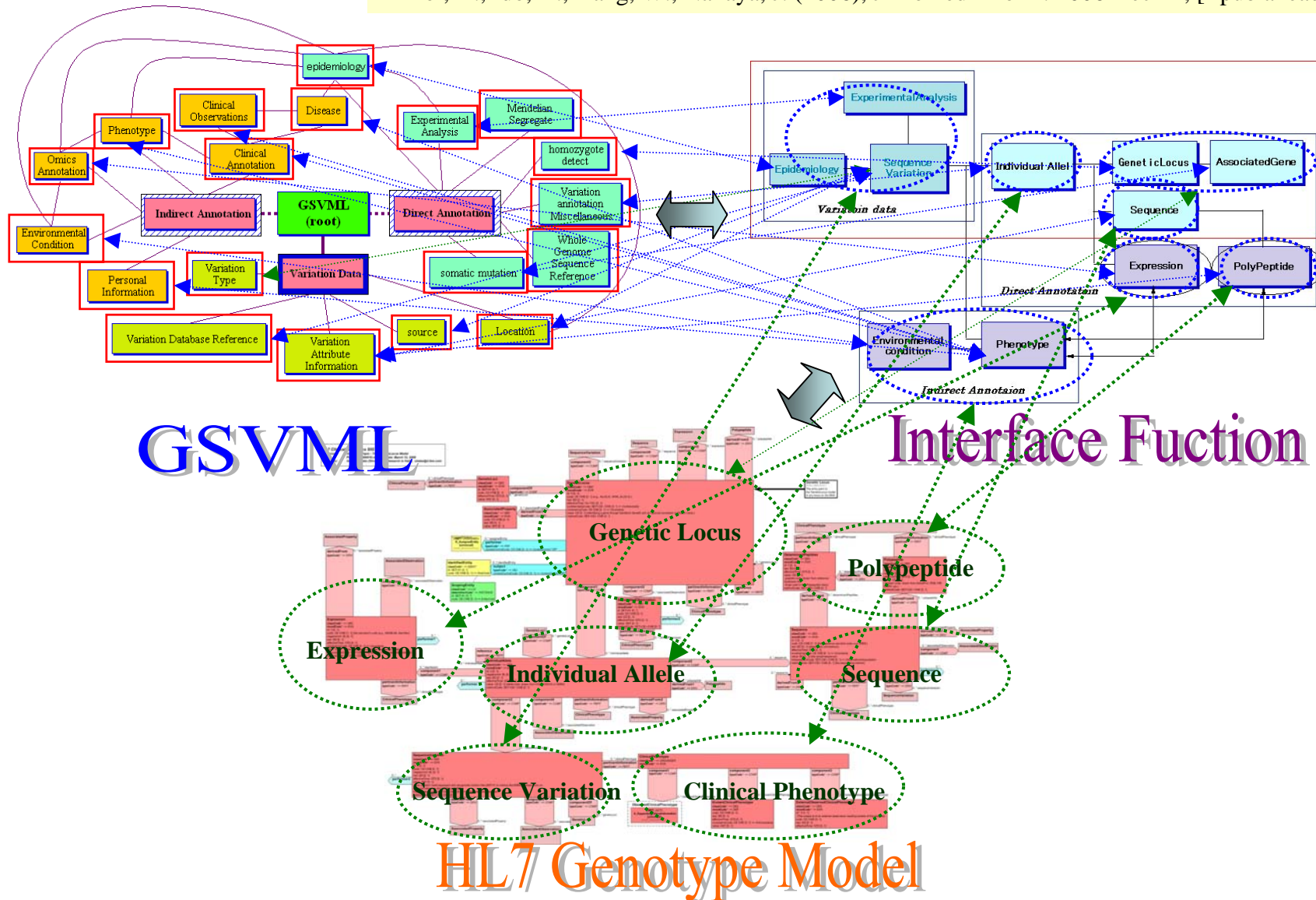
Entry Point is Genetic Locus.



This model packed the DNA sequence variation data, gene expression data, and the clinical phenotype in the relatively plain informational model.

# Interface Model between HL7 Genotype Information Model and GSVML

Hiroi, K., Ido, K., Yang, W., Nakaya, J. (2006), J Biomed Inform. 2006 Dec 24; [Epub ahead of print]





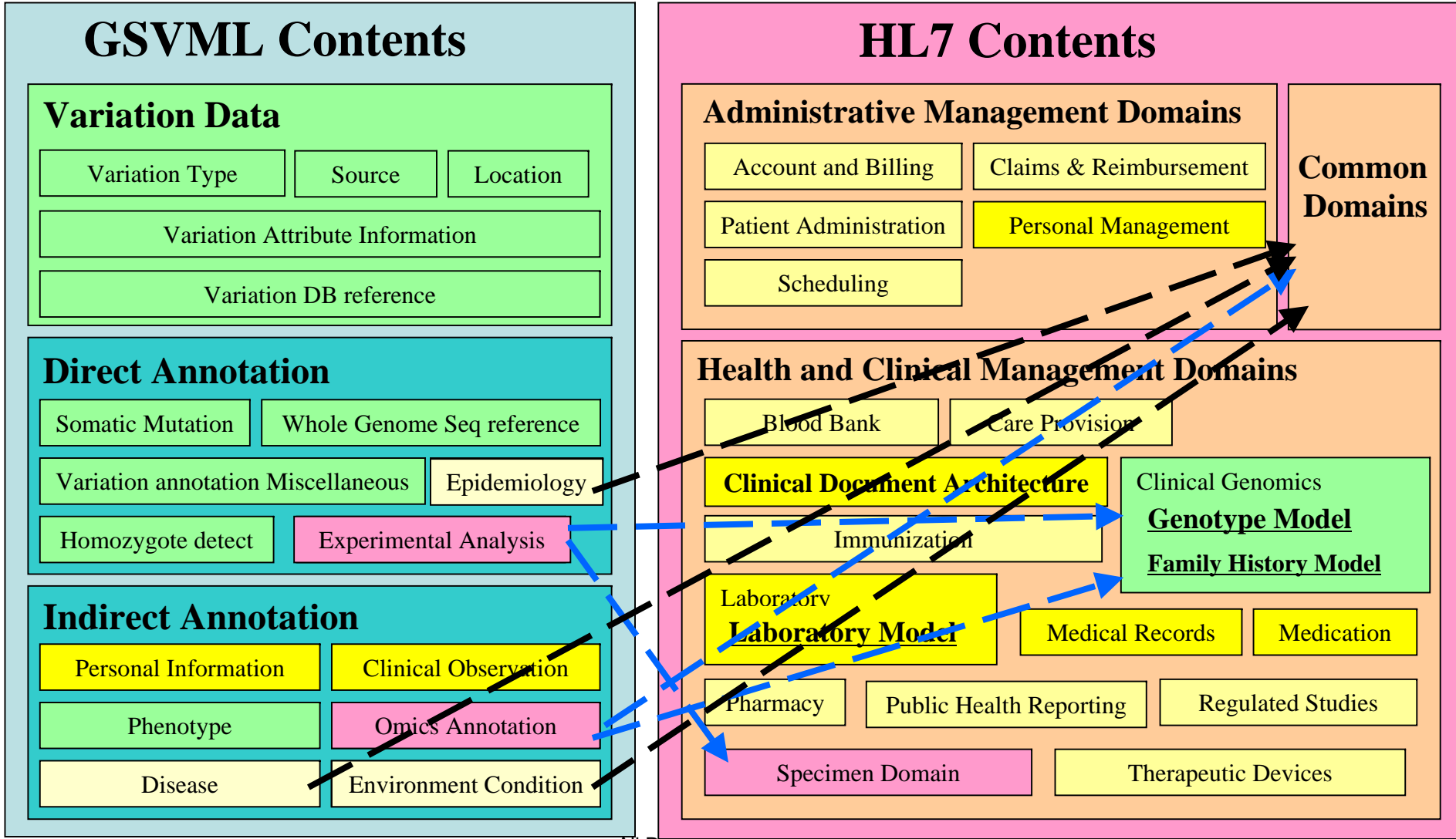
# GSVML contents in total HL7 contents

Already have in HL7GM

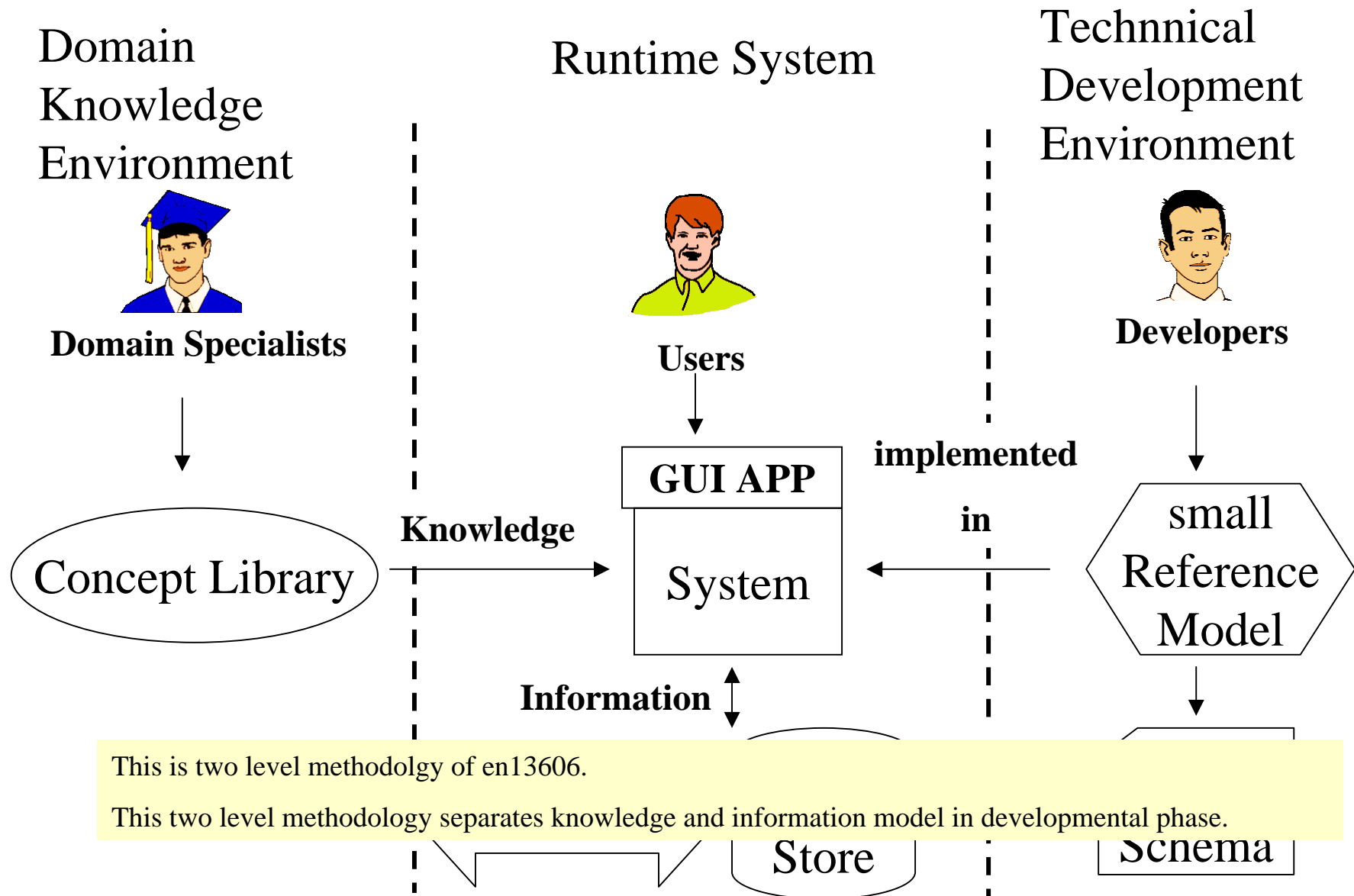
Already have in other HL7 Model

Partially have in HL7

Not have in HL7



# CEN en13606 Two Level Methodology



This is two level methodology of en13606.

This two level methodology separates knowledge and information model in developmental phase.

Already have in HL7GM

Already have in other HL7 Model

Partially have in HL7

Not have in HL7

## GSVML Contents

### Variation Data

Variation Type

Source

Location

Variation Attribute Information

Variation DB reference

### Direct Annotation

Somatic Mutation

Whole Genome Seq reference

Variation annotation Miscellaneous

Epidemiology

Homozygote detect

Experimental Analysis

### Indirect Annotation

Personal Information

Clinical Observation

Phenotype

Omics Annotation

Disease

Environment Condition

## en13606 Contents

Part1  
Reference Model

Part2  
Archetype

Part3  
Reference archetype & term list

Part4  
Security

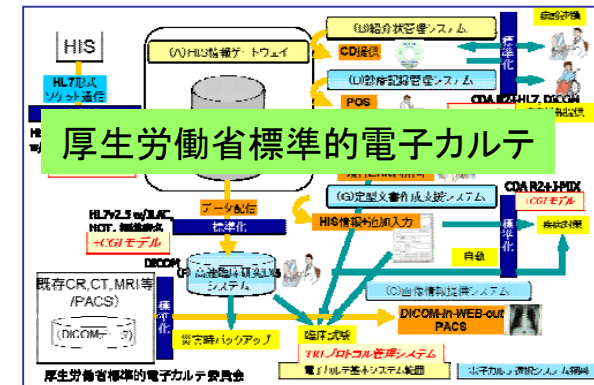
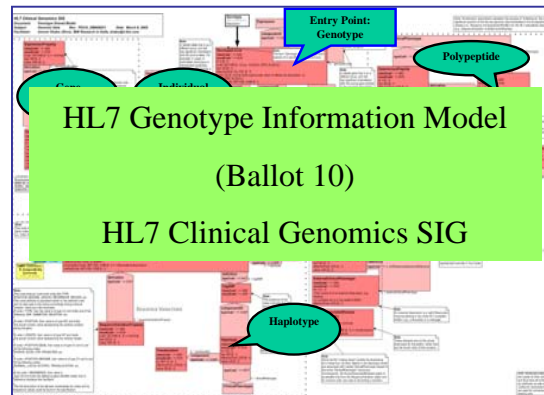
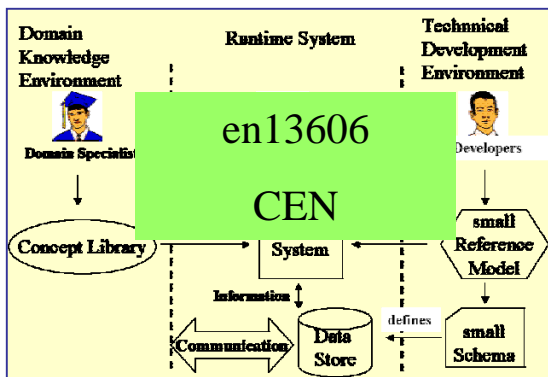
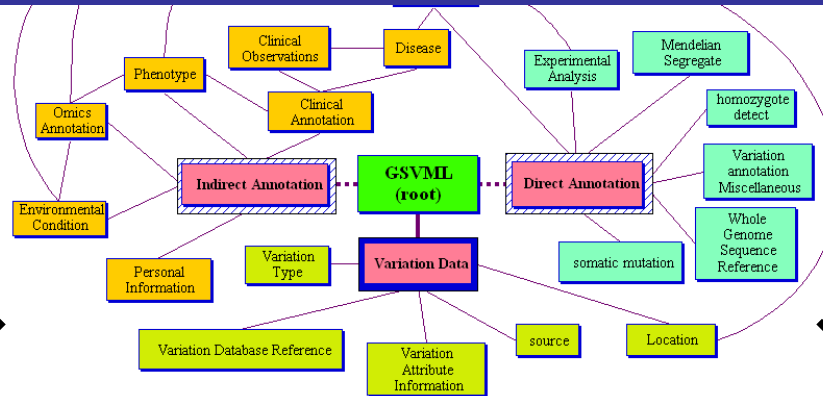
Part5  
Communication Protocol

Part 5 can include all of GSVML contents.  
OR  
Mapping GSVML contents to Part1-4.

# GSVMLと研究用アドオン電子カルテ

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

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



研究用アドオン電子カルテ

国際標準次世代電子カルテ

厚生労働省標準電子カルテ機能追加

# SNOMED Clinical Terms (SNOMED CT)

It is coordinated good.

内容: Clinical Health Care terminology & infrastructure : 臨床、病理、疾患、解剖を含む

各国の国家的EHRプロジェクトで採用

- 英国Npfit (Simplified Conceptual Architecture) プロジェクトで国家機関NHSが購入
- 米国NHII (National Health Information Infrastructure) プロジェクトで国家機関NLMが購入
- カナダ、リトアニア、デンマーク、オーストラリアでも採用

It is expensive, detailed, and complicated.

Atomic Core (要素的な概念単位) の不明瞭性

米国私企業により独占的に販売  
高額

日本の医学体系と合わない面もある

We need more simplified but enough clinical ontology.

- Logical Basis -

Ontological Architecture  
based on  
**Evidence Based Logical Atomism**

各階層において論理的情報単位(論理的原子)を定義し、  
その信頼性を証拠群で評価する

Jun Nakaya, Tetsuo Shimizu: Knowledge Architecture based on  
Evidence Based Logical Atomism for Translational Research,  
International Journal of Computer Science and Network Security,  
vol. 6-2, ISSN:1738-7906, 175-179, 2006

Russel, B., “Philosophy of **Logical atomism**. Monist,” vol. 28,  
495-527, 1918

# Disease Terminology 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

肝臓

肝硬変

胆汁性肝硬変

原発性胆汁性肝硬変 (PBC)

二次性胆汁性肝硬変

特発性肝硬変

心臓性肝硬変

肝腫瘍

原発性肝腫瘍

良性腫瘍

肝血管腫

肝細胞腫

限局性結節性過形成 (FNN)

悪性腫瘍



Anatomical Hierarchy

+

Etiological Hierarchy

肝膿瘍 (liver abscess)

検査所見

画像診断

胸部 X-p

右横

US

cystic pattern

CT

plain CT

low densit

contrast CT

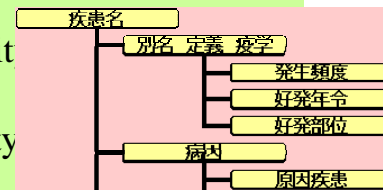
iso density

dynamic CT

基本は箇条書き  
不要な修飾語・語尾を避ける  
～である etc.  
～が所見として見られる etc.  
知識タグとして見出しを利用  
見出しを先に記述する  
その説明は後に記述する  
矢印の活用

Unified Description Rules

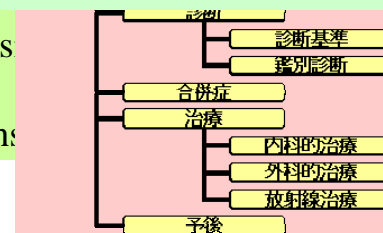
例: 貧血↑, LDH↑ etc.  
全角・半角の区別明確化



Logical Atom:

3rd Normalized Skeleton Template

for Disease



臨床23科の疾患知識 (インターンレベル) を統合

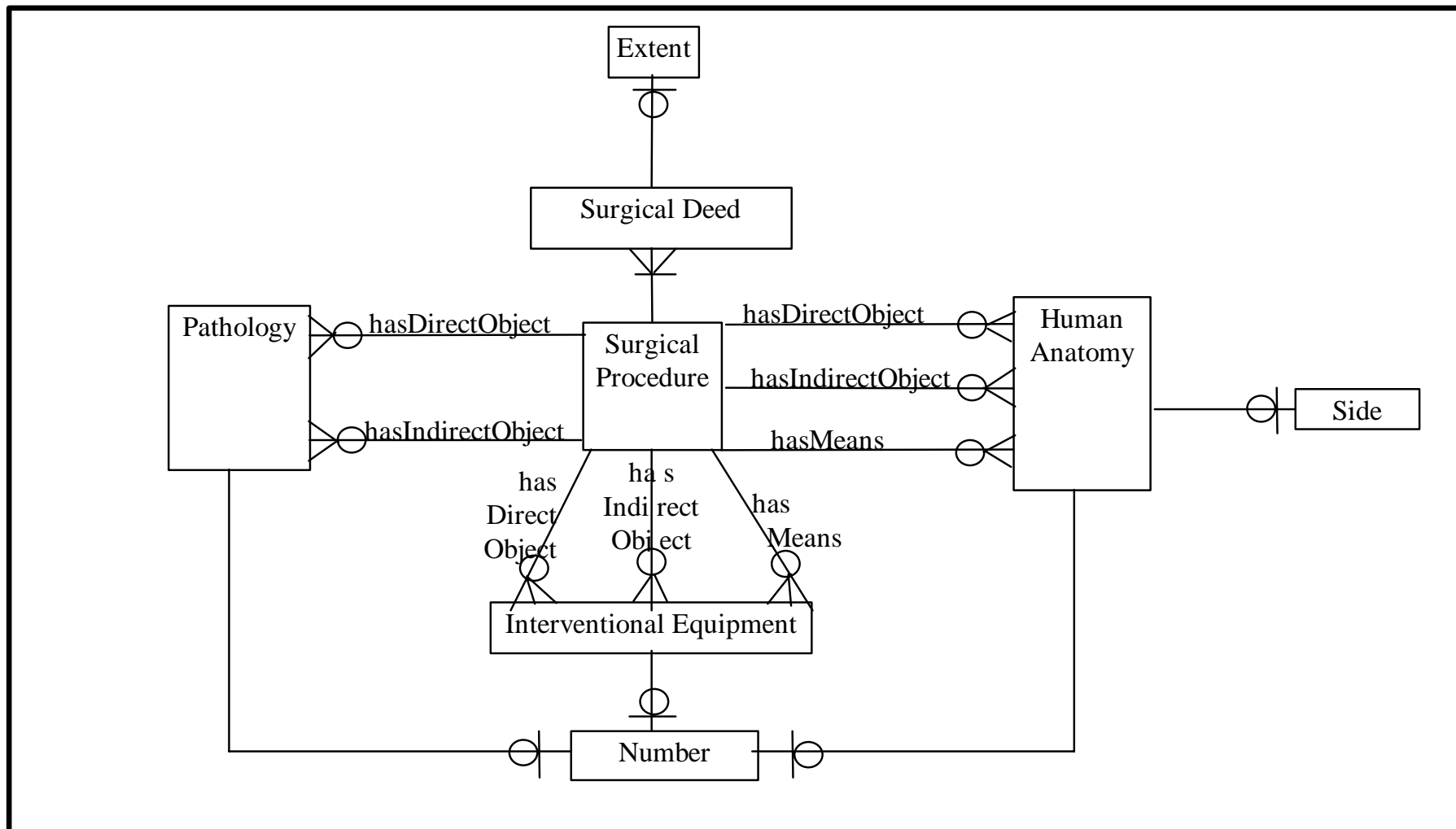
# Comparison Table

GSVML	SNOMED-CT
Clinical annotation	Finding (Clinical Findings), Disease (Clinical Findings)
Disease	Disease (Clinical Findings)
Disease Description	Disease (Clinical Findings)
Name	Disease (Clinical Findings)
Synonym	Disease (Clinical Findings)
Definition Concept	Finding (Clinical Findings)
Classification	Finding (Clinical Findings), Staging and scales
Etiology	
Etiology Description	Finding (Clinical Findings), Body structure, Organism, Substance, Pharmaceutical / biologic product, Context-dependant categories
Etiology Condition	Environments and geographical locations, Social context, Context-dependant categories, Staging and scales
<p>GSVML (LAアーキテクチャー)では、情報の分解、対比、交換が可能            → SNOMED-CTへの変更、マッピングが比較的容易</p>	



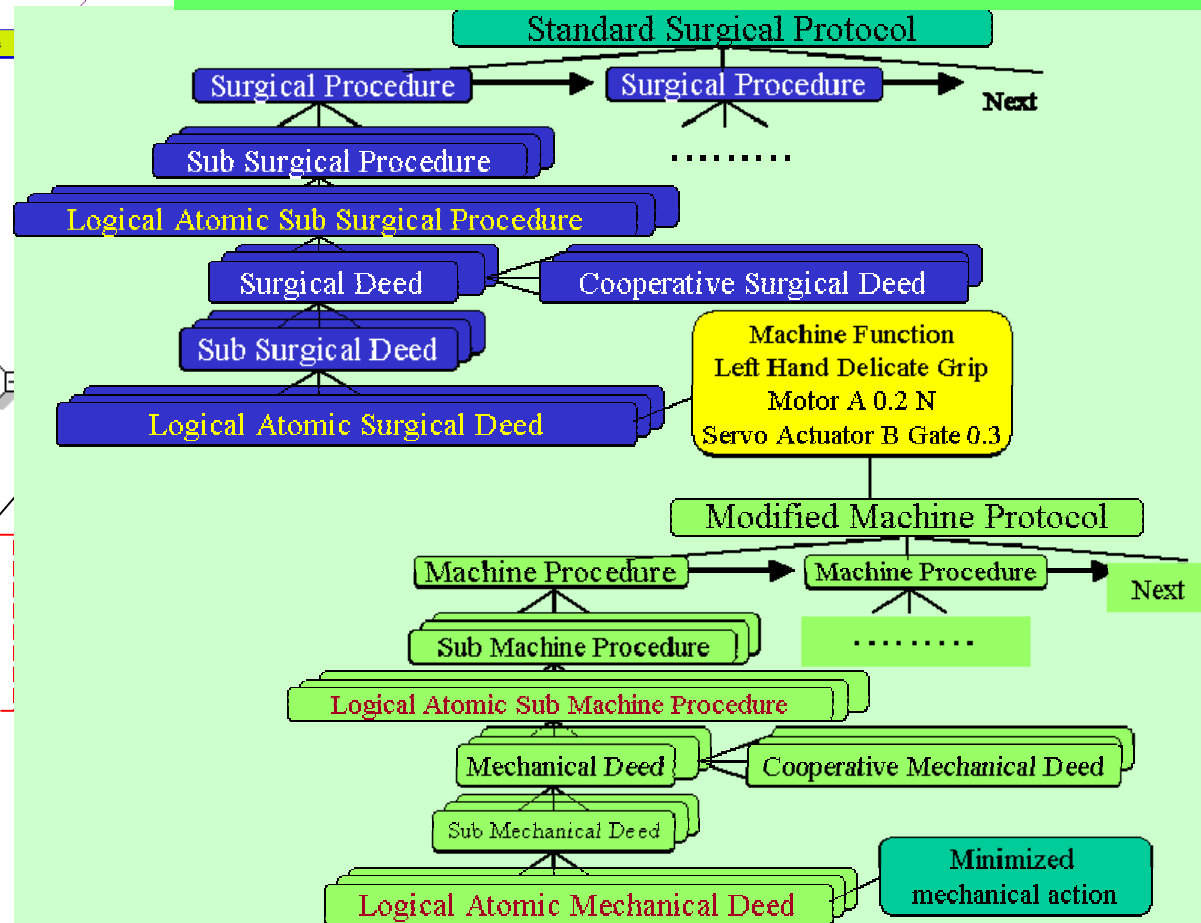
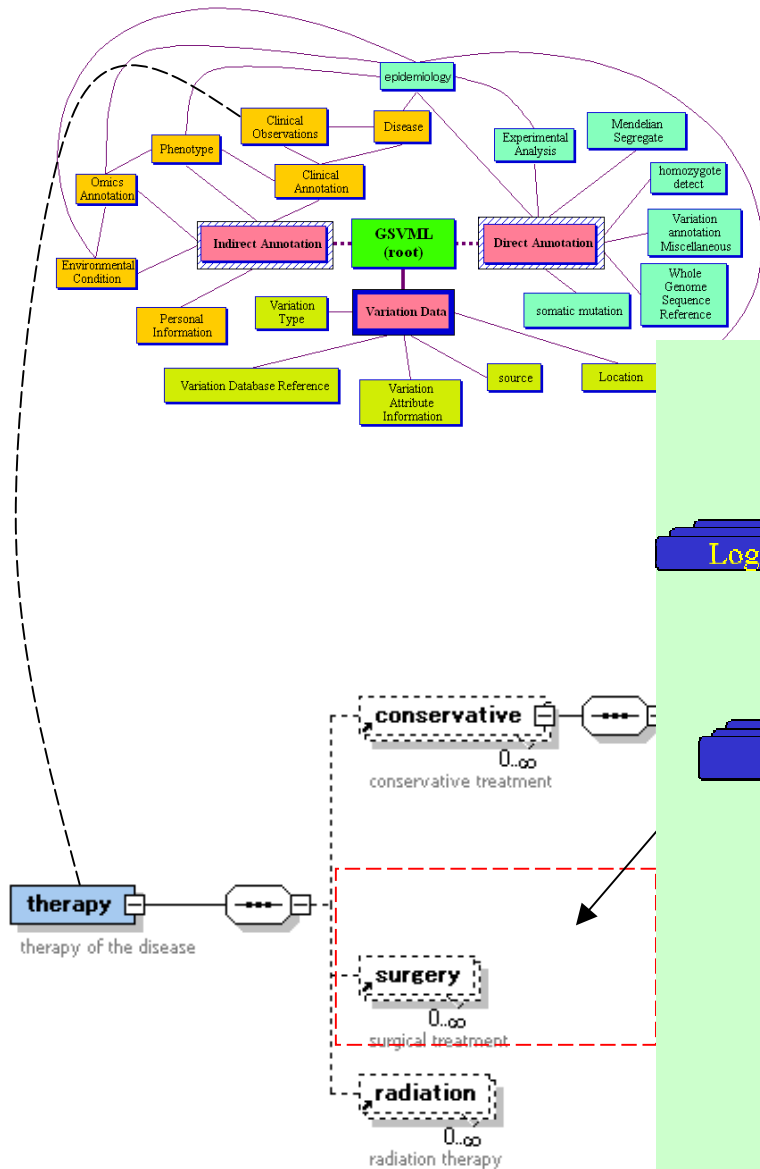
# CEN Model for Surgical Ontology

The CEN model includes the notion of a surgical procedure which is distinct from the individual deeds which it comprises. In CEN, a Surgical Deed is a part of the definition of a Surgical Procedure, and all Surgical Procedures must have one or more Surgical Deeds in their definition.



# Surgical Protocol Attachement for Robot Asisted Surgery

"An intelligent data format for robot-assisted surgery: Challenges and solutions in software control.", Jun Nakaya, Stanford University, Apr. 19 2007



Surgical Ontology based on Logical Atomism

ご清聴ありがとうございました