A Standard-based Approach for Knowledge Representation

MIE – Oslo Norway, Aug 2011

Oral Presentation

Ariel Farkash, Hani Neuvirth, Yaara Goldschmidt, Amnon Shabo – IBM Haifa Research Lab, Israel
Costanza Conti – IMS-Istituto di Management Sanitario SRL, Italy
Federica Rizzi – KOS Genetic SRL, Italy
Stefano Bianchi – Softeco Sismat SRL, Italy
Erika Salvi, Daniele Cusi – Hypergenes Consortium, Fondazione Filaterete, Italy
Some Terminology

- Data – The raw clinical or genomic patient data
- Knowledge – An understanding of the studied disease that is not specific to any patient
  - Knowledge may publically available
  - Knowledge may be generated within the project scope e.g. analysis on data
- Information – a subject-specific analysis result that can be used as a prediction or for decision support purposes
Dealing with data…

**WE USE CDA**
HL7 Clinical Document Architecture (CDA)
Instance Generation

**Data Source**

**Template Model**

**OWL Ontology**

Mapping local Vocabularies

Standard-based Instances (e.g., CDA)

Conform to the Template Model

Instance Generation Engine

CTS

Java API

Representing constraints

Mapping local Vocabularies

Standard-based Instances (e.g., CDA)
Genotype -> phenotype...

INFORMATION & KNOWLEDGE
Hypergenes Data Model

CDA Template
Header
subject id
....
Body
reference to PD
....
clinical & environmental observations

GV Template
subject id
Genomic Obs
Phenotype

Encapsulation

Raw Genomic Data
subject id
HapMap / BSML / MAGE
or Relational schemas optimized for persistency

Pedigree Template
subject id

GV Template
analysis id
Analysis Workflow
Run results

Observed Interpretive
knowledge

Disease Model

EHR
HL7 Clinical Genomics Domain Information Model Adjusted to Knowledge Representation

Genetic Locus

Genome

Phenotype

Allele

Genetic Loci

Non-locus specific data

Sequence

Peptide

Expression

Sequence Variation
Analysis Knowledge GV Template

- **Analysis meta information**
  - Analysis title and unique identifier
  - Analysis type
  - Performer details: name, organization, OID...

- **Analysis workflow**

- **Analysis results**
  - Effective time of analysis results
  - Risk Allele
    - SNP – Gene id, chromosomal coordinate
    - Risk allele pvalue
    - Non risk counterpart allele
    - Associated phenotype
Subject Specific GV Knowledge Template

- Metadata
  - Effective creation time
  - Unique instance identifier

- Subject details
  - Identification / Organization

- Subject SNPs
  - Genotyping center / device
  - SNP alleles from genomic data
  - Only intersection of SNPs relevant for analysis are taken

- Subject Clinical data
  - HL7 references to observations from clinical records (CDAs)

- Analyses results impact on subject
  - Analysis id / reference to CG instance
  - Risk alleles for subject
The End

Thank You 😊