
Conceptual Modeling of Genomics, Association Studies, and Pharmacogenetics

Xiaohua Zhou, Il-Yeol Song

College of Information Science & Technology, Drexel University

{Xiaohua.Zhou,Song}@drexel.edu

Motivation

- Short-term Goal

- Understand fundamental relationship among genome, genotypes, phenotypes and environment.

- Long-term Goal

- Develop a system that supports biomedical research activities that use both clinical and biomedical data to improve cancer patient care by incorporating the roles of environment, life style and genetics.

Information Model

■ Term Definitions

- Genetic Association Studies examine relationships between genetic variation and disease susceptibility.
- Pharmacogenetics studies the drug responses against genetic variation.

■ Complexity of Data Model Required

- Association studies and Pharmacogenetics demand a variety of information such as genotype, clinical observations, disease development history, demographical information, life style, and living environment.

Information Model (Cont.)

■ Conceptual Modeling

- Correct and informative modeling of these data is very important for bioinformaticians because the model affects data manipulation and the types of queries they can ask as well as the performance of the implemented system.

■ Challenges in Conceptual Modeling

- Intuitive representation of a complex model.
- Changing information needs for representation and query processing
- System performance.

System and Methods

■ The System

- Provides experimental data and knowledge base for genetic studies and Pharmacogenetics, and incorporates other related public resources

■ Methods

- Use package diagram and class diagram of Unified Modeling Language (UML) to represent the system.
- UML: industry standard for object-oriented analysis and design.

System and Methods (Cont.)

■ Methods

- Use package diagrams to show the subsystems and inter-subsystem dependencies.
- Use class diagrams including class name, important attributes and relationships like association, aggregation, generalization and realization to show domain classes and inter-class relationships.
- 1 package diagram and 10 class diagrams in total.

System Package Diagram

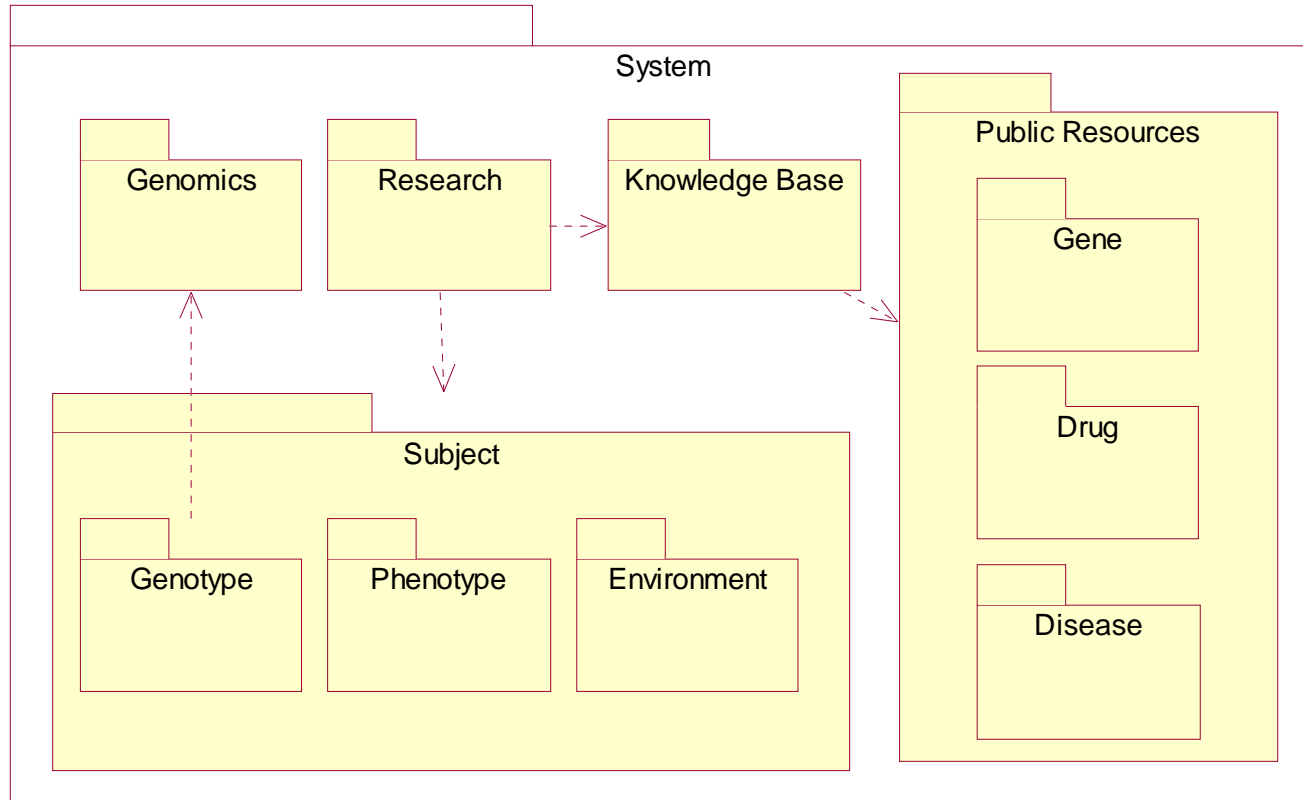


Figure 1 System Package Diagram

Human Genome Model

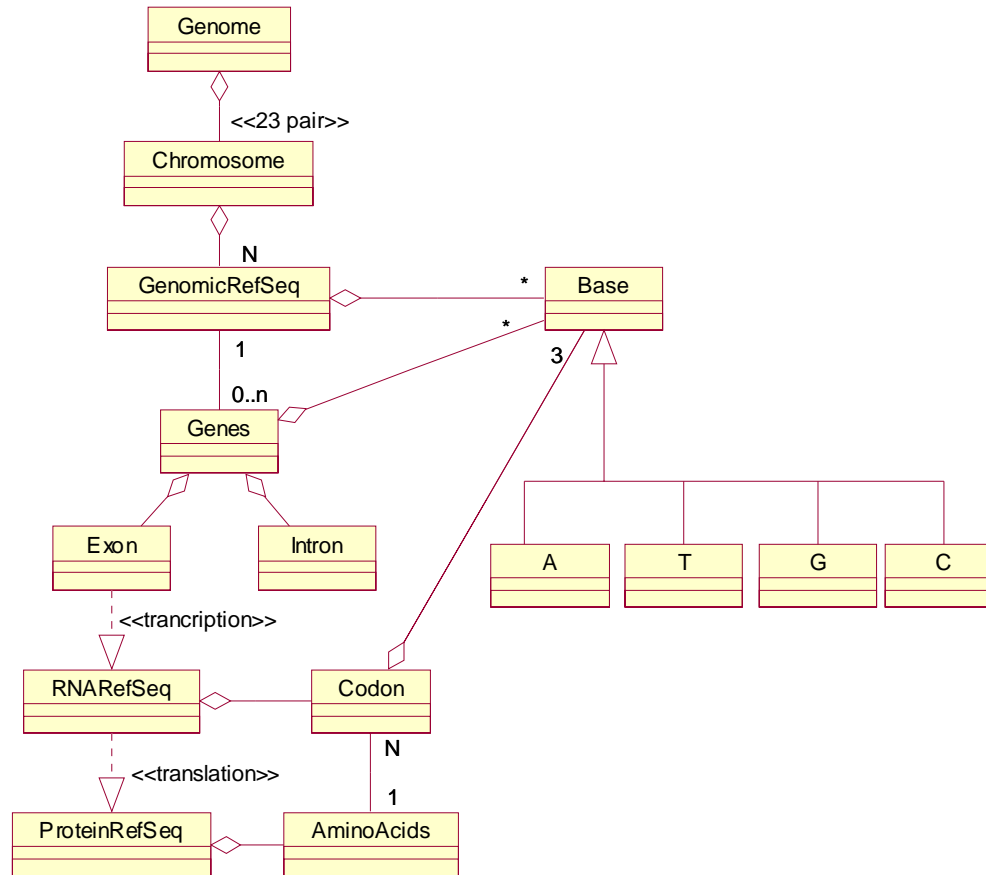


Figure 2. Human Genome Diagram

Subject Information Model

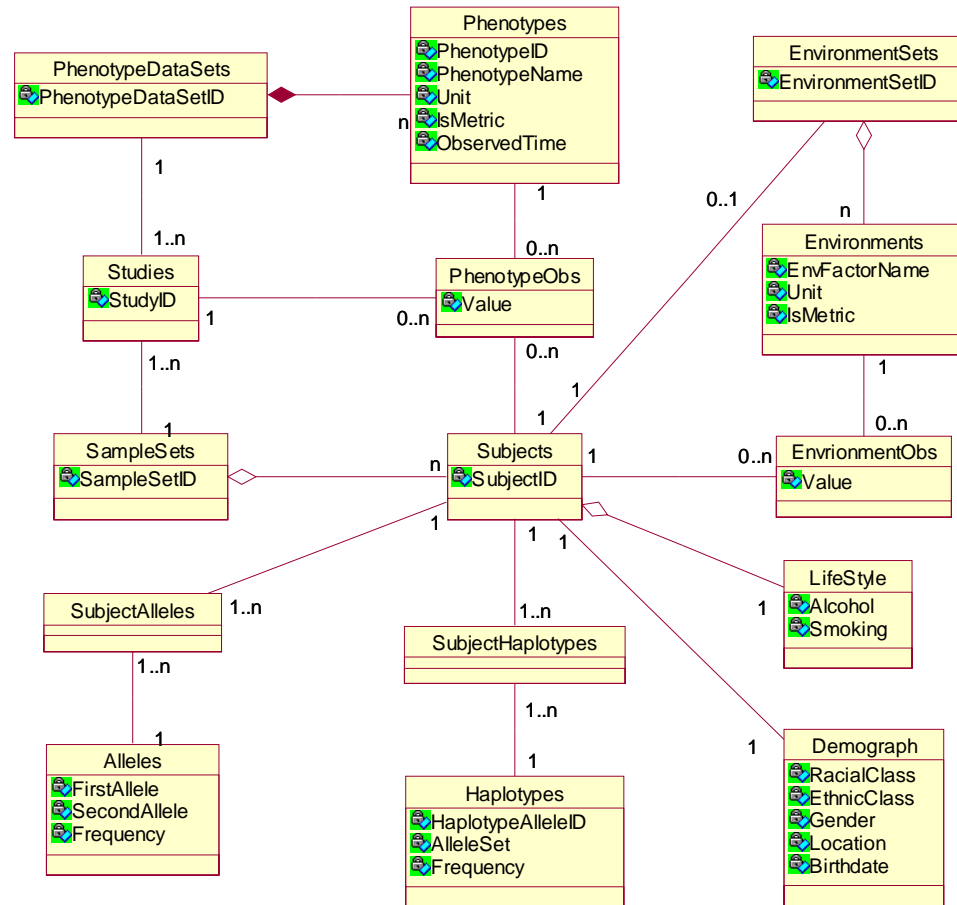


Figure 3. Subject Information Diagram

Research Model

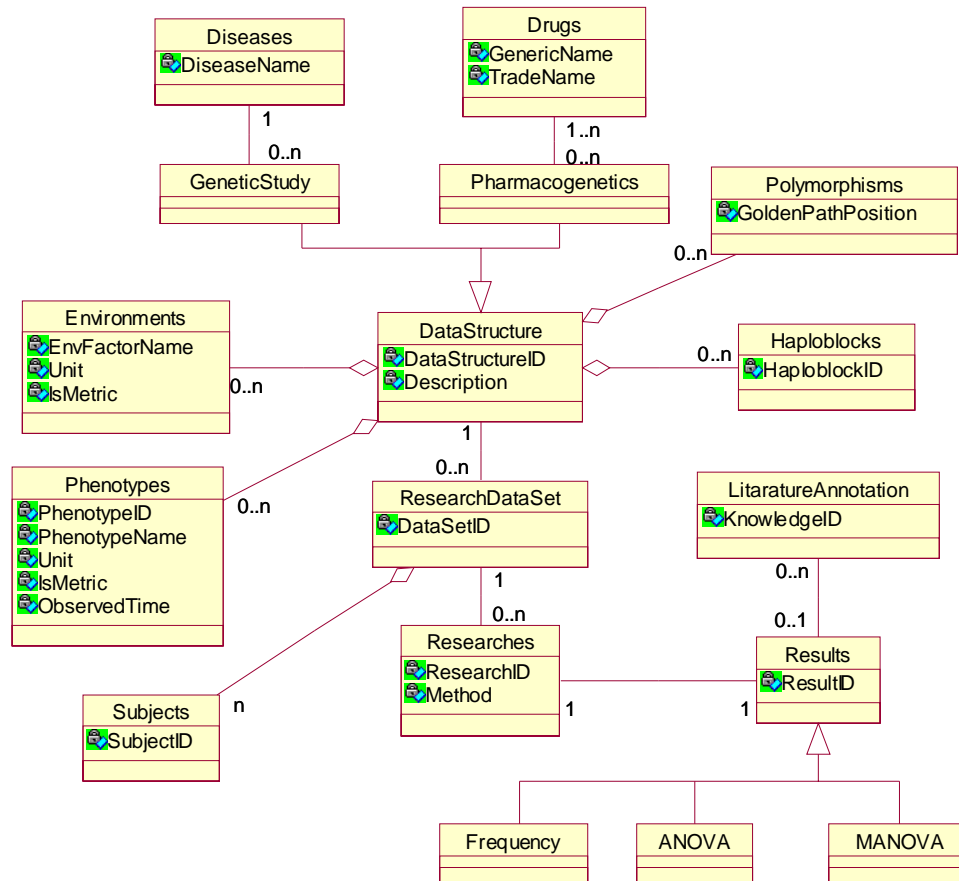


Figure 4. Research Diagram

Genotype Model

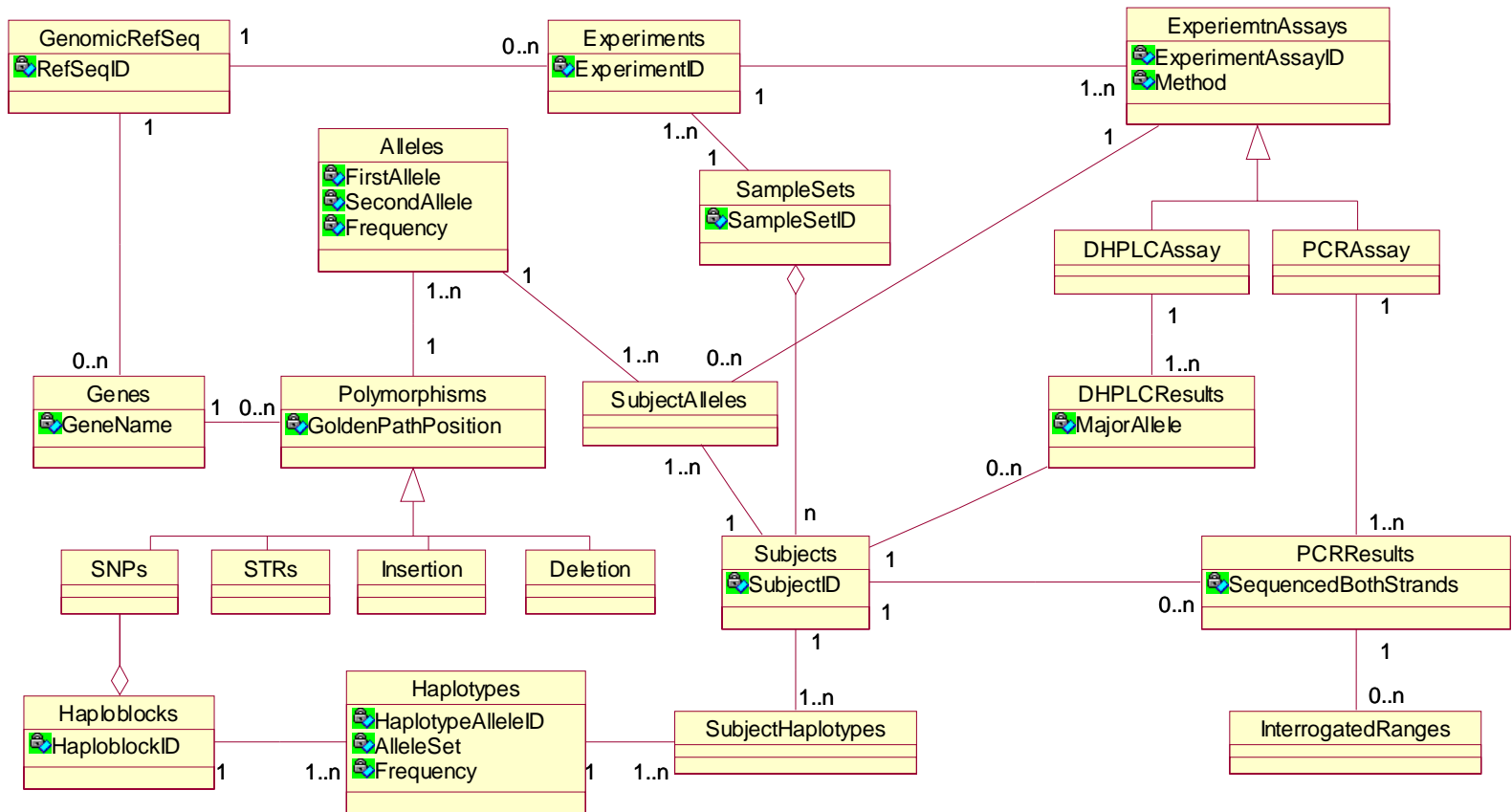


Figure 5. Genotype and Experiment Assays

Phenotype Model

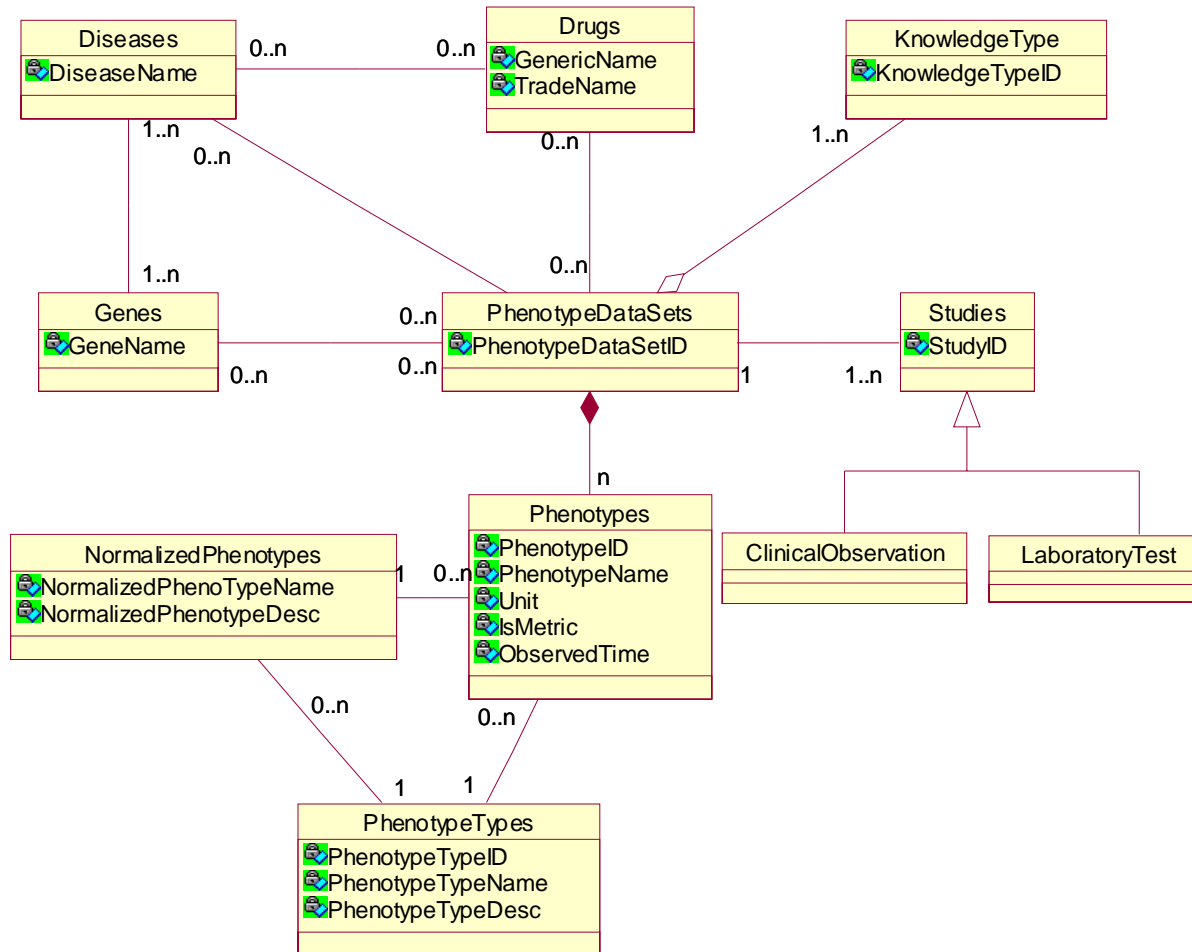


Figure 6. Phenotype Diagram

Models for Public Resources (Gene)

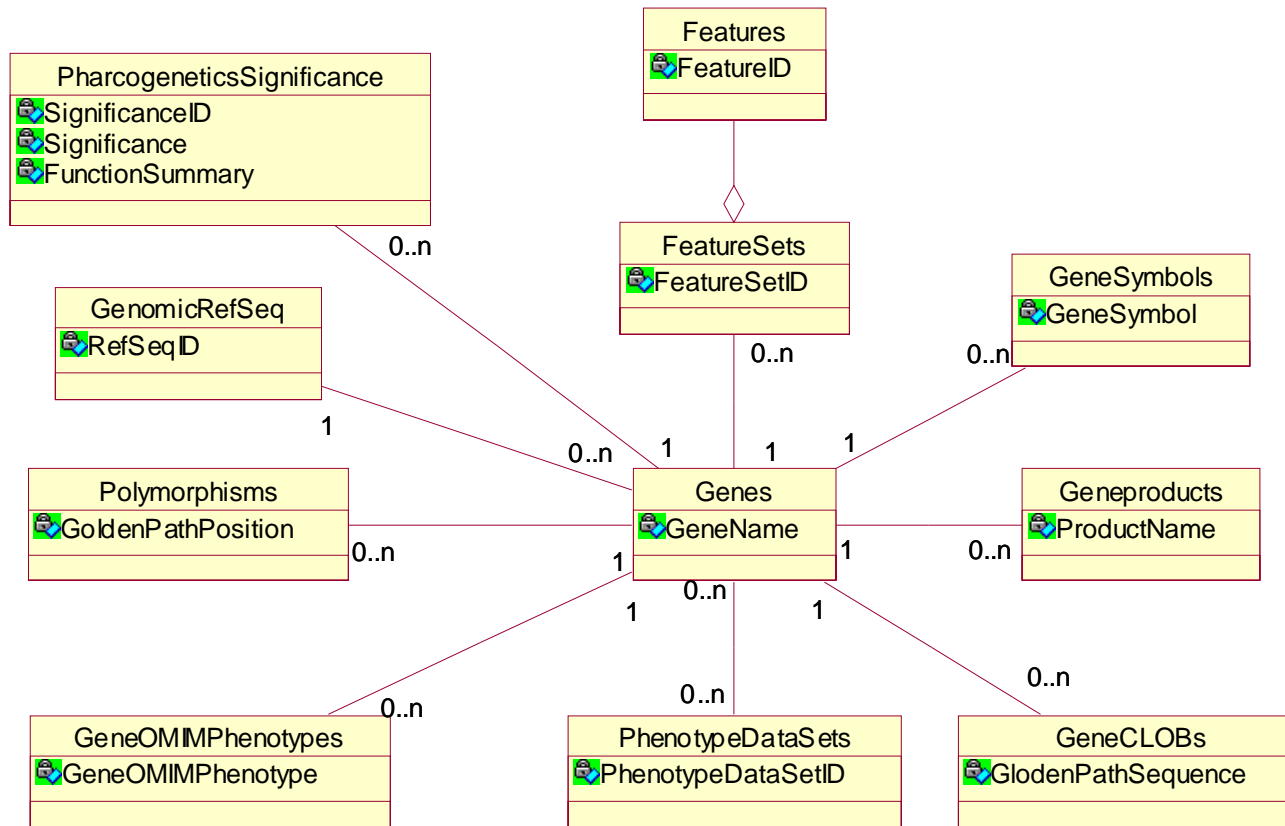


Figure 7. Gene Diagram

Models for Public Resources (Disease & Drug)

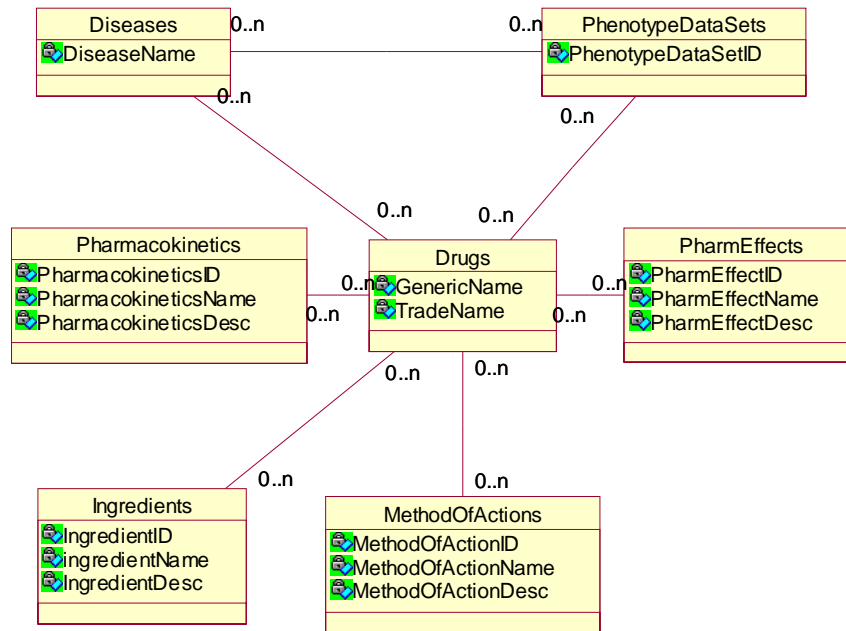


Figure 8. Drug Diagram

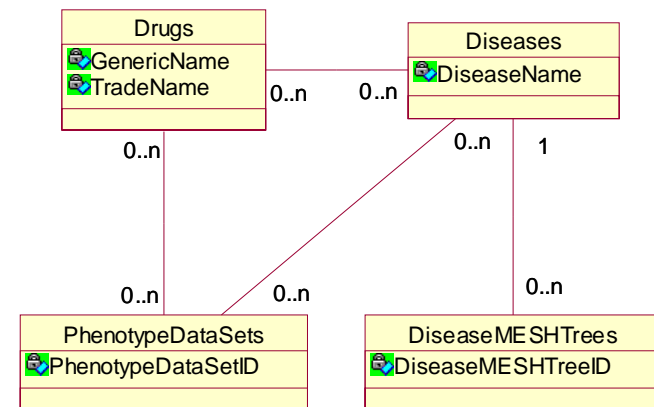


Figure 9. Disease Diagram

Model for Knowledge Base

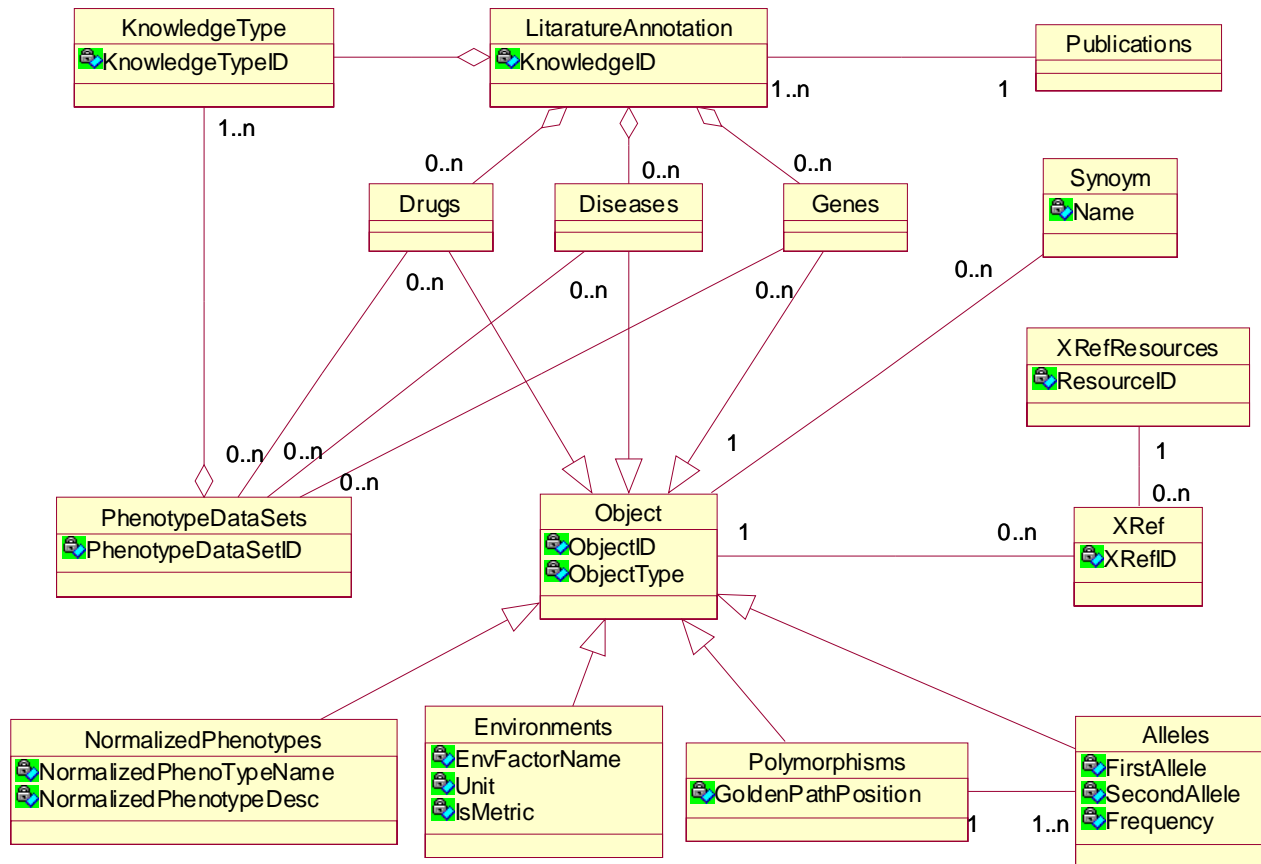


Figure 10. Knowledge Representation Diagram

Concluding Remarks

■ Current Model

- Intuitively represent a complete system for genetic studies and Pharmacogenetics using UML notation.
- Suppose to meet the basic data needs of above mentioned two lines of researches.

■ Future Work

- Implement and evolve
 - Test and meet the users' information needs.
 - Optimize the performance of the system.
- Enhance support for recording of experiment procedures.
- Enhance support for longitudinal studies.