

# **HL7 Clinical Genomics – Domain Information Model**

*The HL7 Clinical Genomics Work Group*

*Prepared by Amnon Shabo (Shvo), PhD*

***HL7 Clinical Genomics WG***

***Co-chair and Modeling Facilitator***

***Pedigree (family health history) Co-editor***

***GTR Implementation Guide Primary Editor***

***HL7 Structured Documents WG***

***CDA R2 Co-editor***

***CCD Implementation Guide Co-editor***

# Overview of Activities

## Main efforts: (chronologically ordered)

### v3:

- Family History (Pedigree) Topic
- **Deprecated:**
  - Genetic Locus / Loci
  - Genetic Variations Topic
  - Gene Expression Topic

### v2:

#### v2 Implementation Guides

\* The IG “Genetic Test Result Reporting to EHR” is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1

### CDA:

- A CDA Implementation Guide for Genetic Testing Reports (GTR)

## Under Development:

### FHIR:

- Genomic resources
- Genetic family history profile

### Common:

- Domain Analysis Models
- Domain Information Model (DIM) describing the common semantics
- DIM is standards-independent; *Semantic alignment among the various specs*

-  Normative
-  DSTU
-  Informative

# CG Domain Information Models (DIMs)

- Represent agreed-upon and common CG semantics
- Are aligned with the CG DAM requirements and models
- **A focal point to align the various CG specifications**
- **Balloted in 9/2014 at informative level**
  - Consisted of two models:
    - Clinical Genomics Statement (CGS)
      - A specialization of the HL7 Clinical Statement model
      - An extension of the GTR Clinical Genomics Statement
    - Family Health History (utilizing the CGS)
  - Passed, as follows: Aff. : 29, Neg.: 12, Abst.: 69, NV: 24
  - All comments regarding CGS - reconciled (need to reconcile Pedigree)

# Clinical Genomic Statement - Core Principles

- **Explicit representation of a 'genotype-phenotype' association**
  - The association should not be pre-coordinated, that is, a phenotype is not represented as part of some genetic variation representation, e.g., a genetic biomarker is an explicit association of two distinct observations and there is no pre-coordination of the interpretation into a single biomarker code
- **The association is a many-to-many relationship, e.g., a genetic variation can be associated with several phenotypes and a phenotype could be associated with multiple 'genotypes'**
- **It is possible to populate merely the genotype part, thus serving genetic variation representation in general, even if no phenotypic data is available**

# Genetic Observations

- Capture various types of data, from known mutations to sequencing-based variations (e.g., somatic mutations in tumor tissues) to structural variations (e.g., large deletions, cytogenetics) to gene expression and so forth
- The core class `GeneticObservation` contains data that is not the result of downstream analysis, i.e., that cannot be derived analytically from another observation available to performer
- Link to `GeneticObservationComponent` class (zero to many) enables the post-coordination of any core data item (e.g., HGVS string) to its unambiguous basic primitives
- Enabling encapsulation of key raw omics data that is the basis of creating this clinical genomics statement
- The encapsulated data can have links to the full-blown omics data (e.g., NGS sequencing data)

# Phenotypes

- **Phenotype class is abstract and must be instantiated**
- **At instantiation time, distinguish between interpretive phenotype and observed**
- **Interpretation is represented as a distinct observation with its own attributes such as time, method, performer, etc.**
- **Links to knowledge sources / methods / services used for the interpretation**

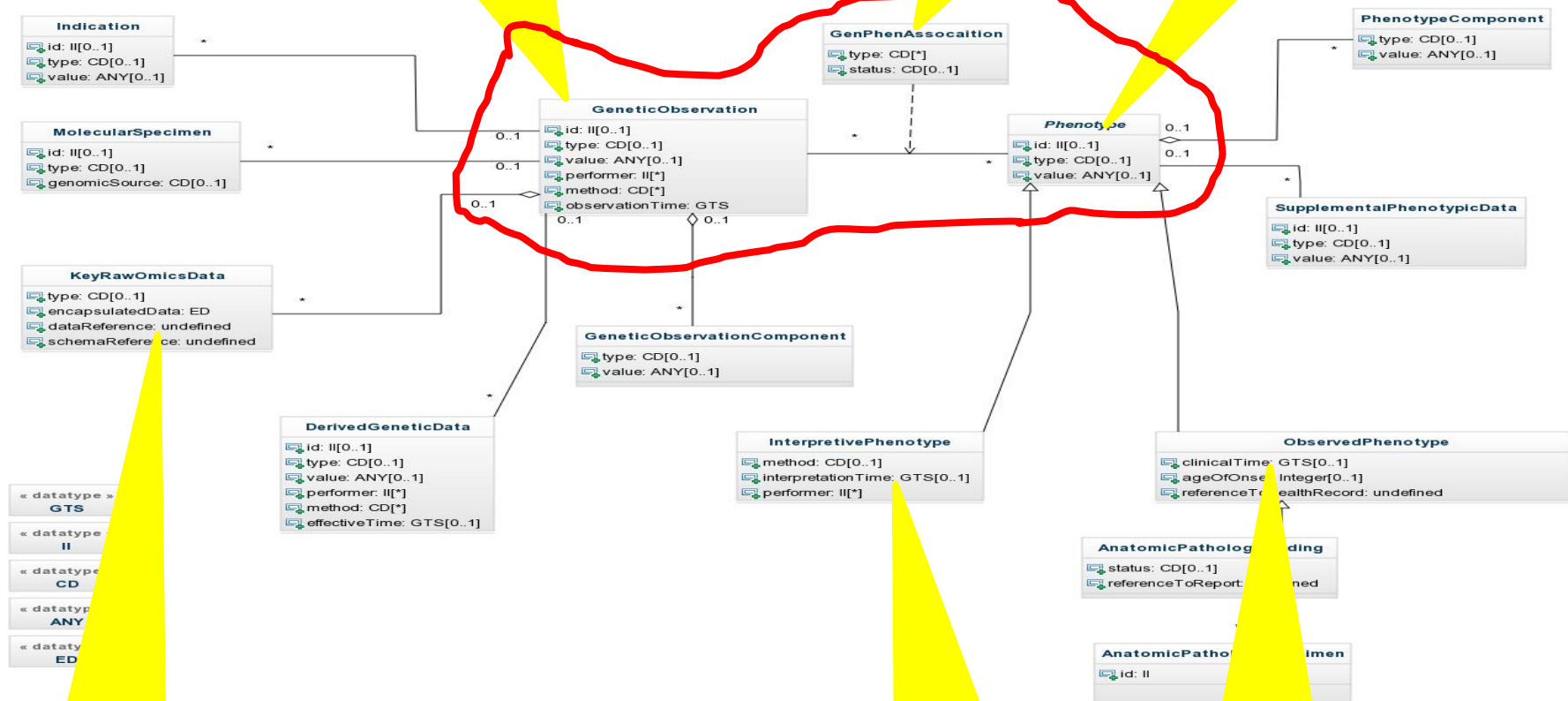
# CG DIM Overview

The backbone:

Gen-Phen association

The genetic observation(s)

The Phenotype(s)



Encapsulated raw data

Interpretive phenotype

Observed phenotype

# **Class: GeneticObservation**

- **Representation of any observation that cannot be analytically derived from other observations made by the performer, including genetic loci in the size of a gene, e.g., genes, biomarkers, point DNA mutations, etc., as well as structural variations, e.g., large DNA deletions or duplications, etc.**
- **Note that omics data derived from this observations, e.g., RNA, proteins, expression, etc., should be represented using the linked DerivedGeneticData class.**



# Class: Phenotype (abstract)

## ■ Description:

- Representation of any data considered phenotype of the source genetic observation. For each genetic observation (e.g., variation), there could be multiple phenotypes, however, each phenotype object should represent a single phenotype. Defining a set of properties to be a 'single' phenotype might be determined by the intended utility of this phenotype, e.g. in patient care or in clinical trials or earlier on in different phases of research.

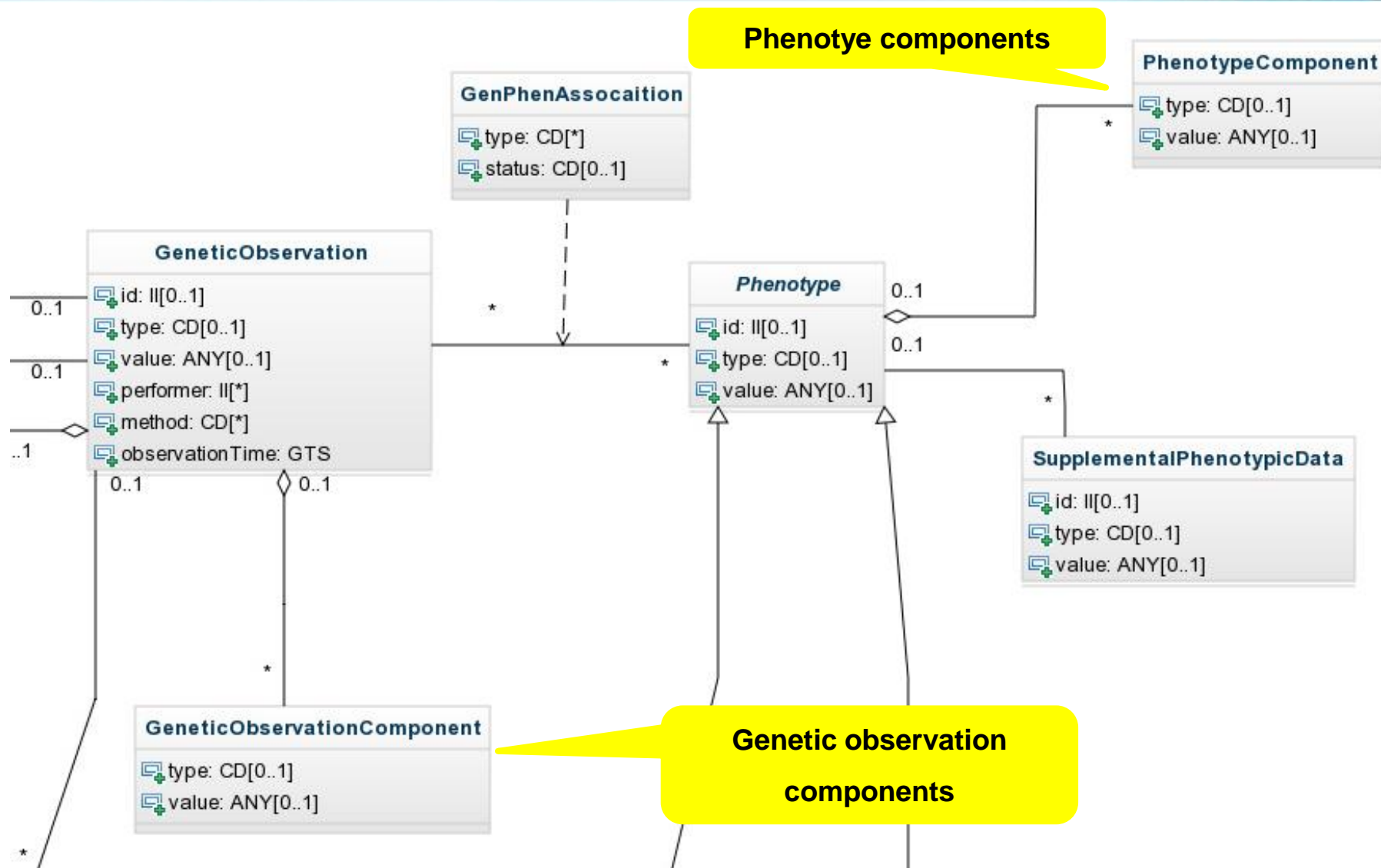
## ■ Attributes:

- id: a globally-unique id of this observation
- type: the type of phenotypic observation, e.g., drug responsiveness
- value: the actual observation (should be drawn from dedicated ontologies, e.g., Human Phenotype Ontology, <http://www.human-phenotype-ontology.org/>)

# Association Class: Gen-Phen

- **This association class adds characteristics that cannot be represented through either side of the core association, e.g.,**
  - **type of relation like causality (pathogenic if phenotype is disease) or non-causative (benign)**
  - **status of this relation, e.g., established, provisional, presumed, preliminary, etc."**

# Post-coordination of Gen. and Phen. Obs.



# **Class: GeneticObservationComponent**

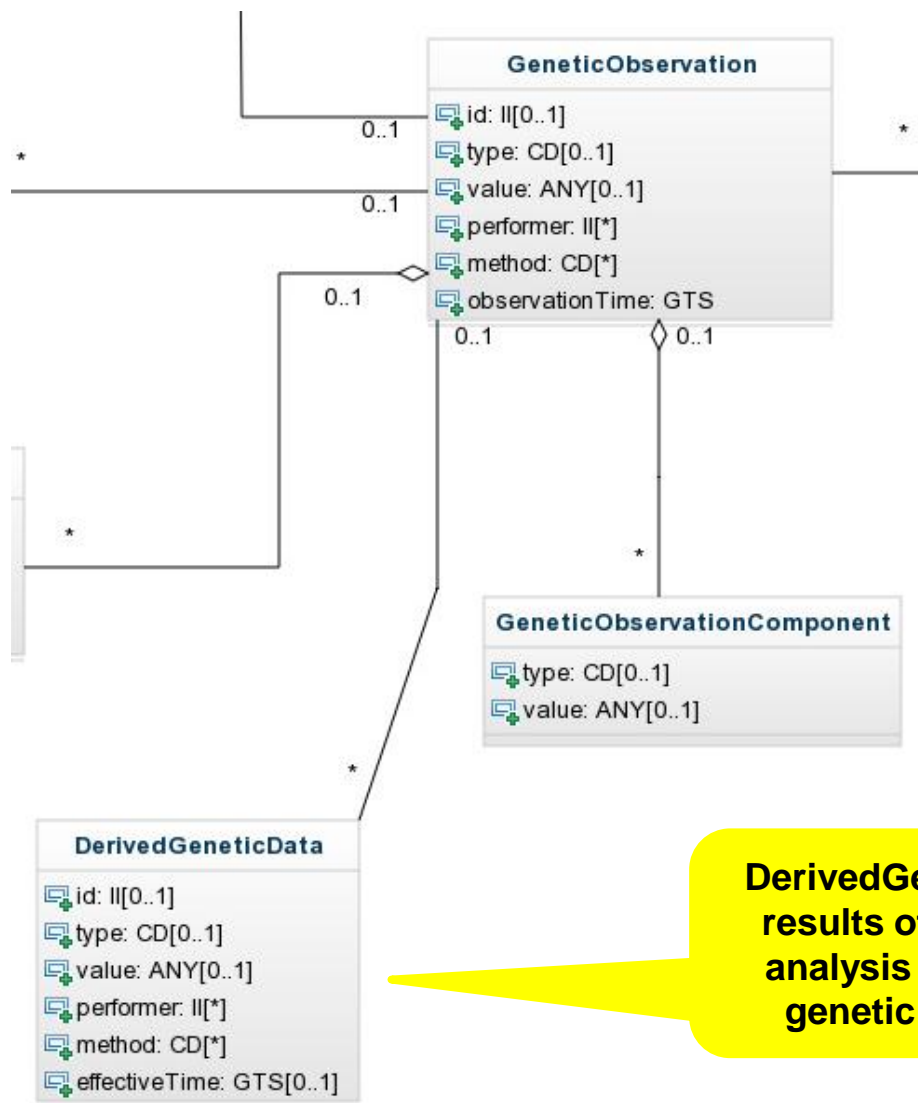
- **Description:**

- Representation of components of the core genetic observation in a post-coordinated manner, i.e., breaking it down to its primitives in order to disambiguate data held by the GeneticObservation class, e.g., an HGVS string.

- **Attributes:**

- **type:** the type of genetic observation component, e.g., variation location, variation length, etc.
- **value:** the actual component observation, e.g., if type is variation base length then value is the actual length

# Derived Genetic Data



**DerivedGeneticData: the results of downstream analysis of the source genetic observation**

# **Class DerivedGeneticData**

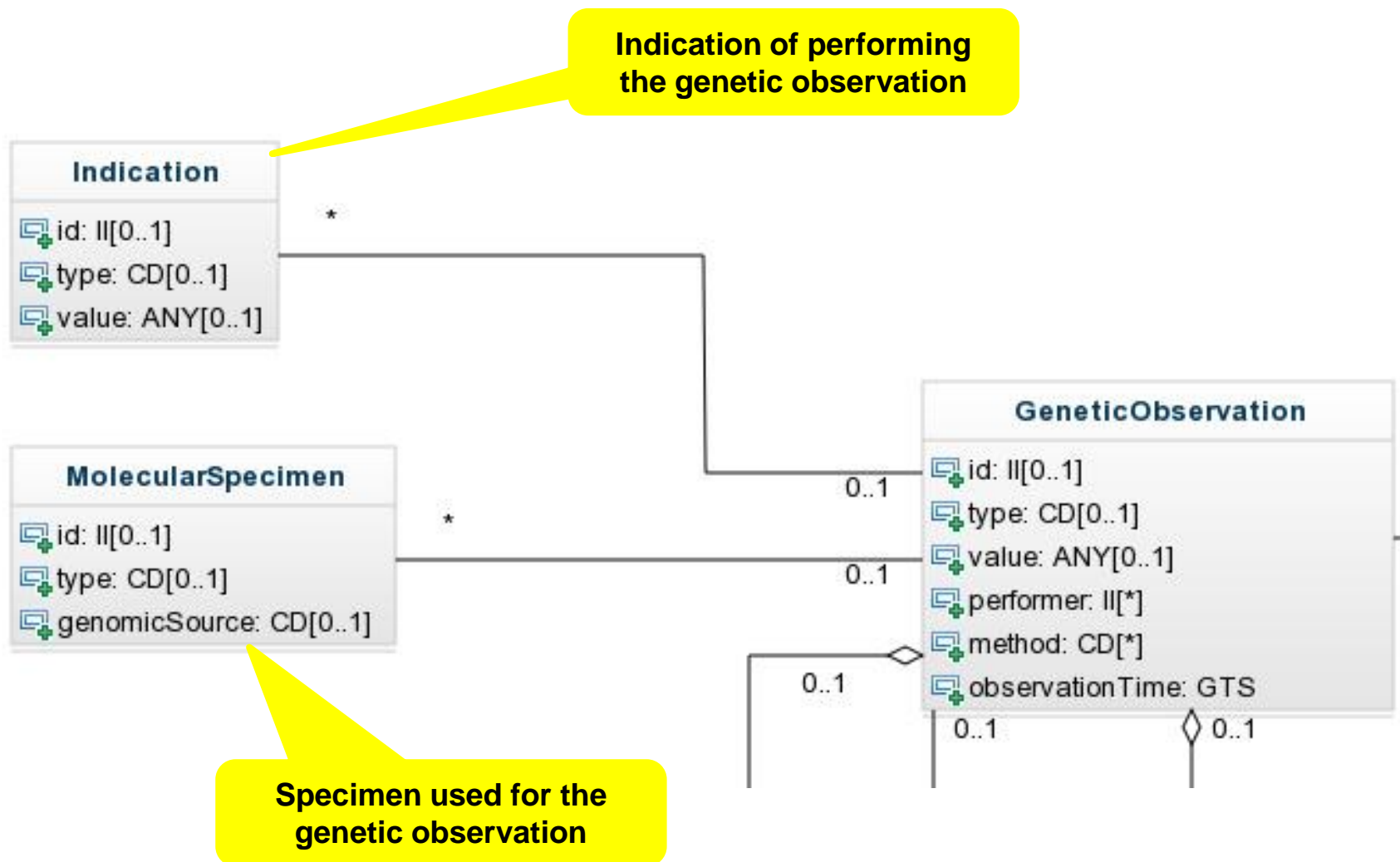
- **Description:**

- Representing any downstream data derived from the core genetic observation (including omics data), e.g., RNA, proteins, expression, etc.

- **Attributes:**

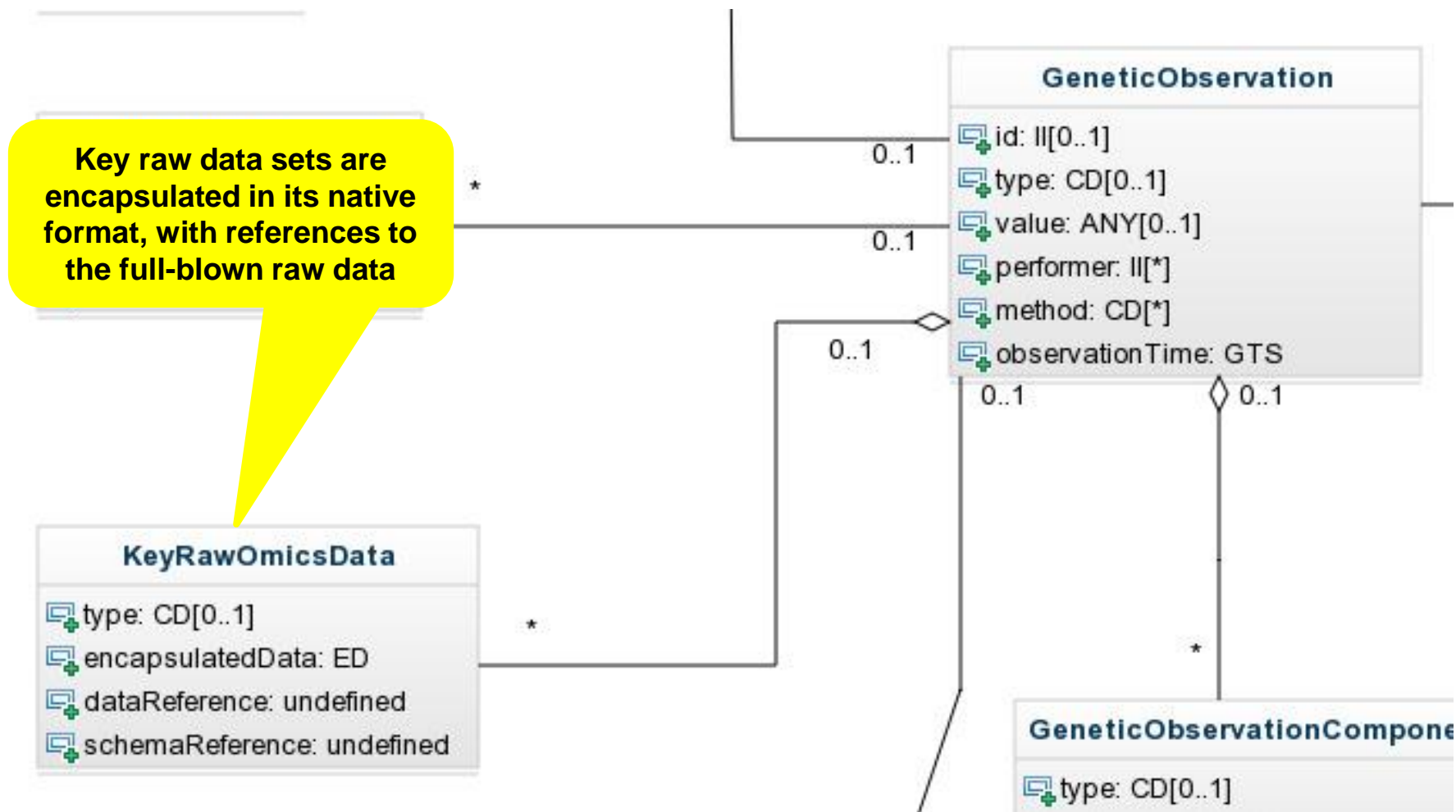
- **id:** a globally-unique id of this observation
- **type:** the type of observation, e.g., characteristics of the resulting protein
- **value:** the actual observation
- **performer:** e.g., unique id of the genetic lab that generated this data

# Molecular Specimen and Indication





# Encapsulated Key Raw Data Sets





# Class: KeyRawOmicsData

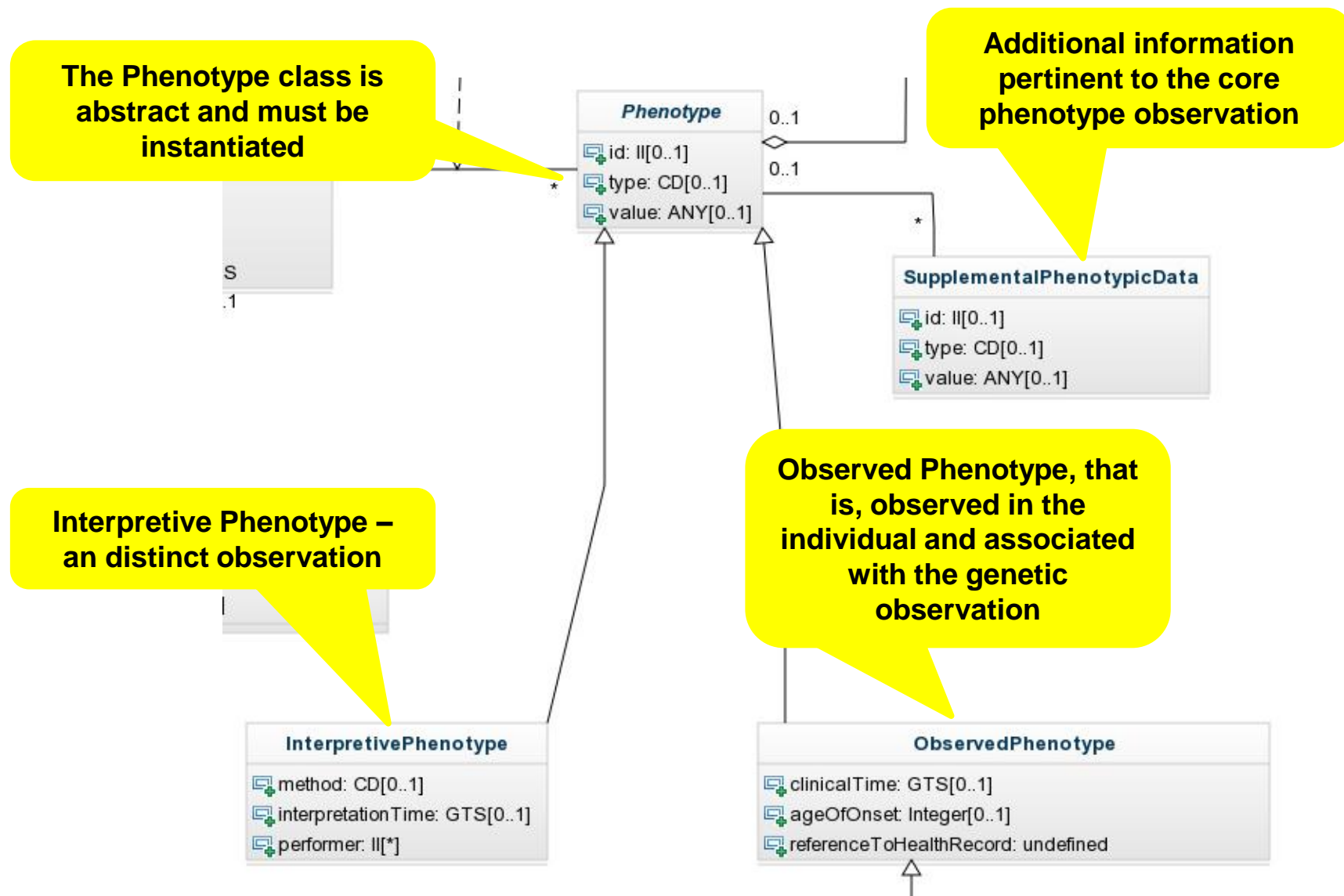
## ■ Description:

- Representation of key data extracted from mass raw genomic data, e.g., an exon's sequence that is considered key to this clinical genomics statement, is extracted from some larger sequencing (up to whole exome sequencing) and encapsulated within this statement structure. Another example: certain lines of a VCF file (preferably its 'clinical-grade' version). The encapsulated data should contain references to the full-blown raw omics data.

## ■ Attributes:

- **encapsulatedData**: a placeholder for the key raw omics data, e.g., DNA sequences, variants calls, expression data, etc. Could be any type of data encoding: binary, text, XML, etc.
- **reference**: a pointer to the repository containing the full-blown data sets out of which these key data items were extracted from
- **type**: designating the notation used in the encapsulatedData attribute, e.g., VCF, FASTA, etc.
- **schemaReference**: a pointer to the schema governing the representation of the data in the encapsulatedData attribute

# Interpretive versus Observed Phenotype



# **Class ObservedPhenotype (ext. *Phenotype*)**

## ■ **Description:**

- This is a sub-class of **Phenotype**, representing a phenotype observed in the subject, e.g., if the somatic mutation is known to be the reason of positive responsiveness to some drug, then that responsiveness is an observed phenotype.

## ■ **Attributes:**

- **clinicalTime**: the effective time for the subject of this observation
- **ageOfOnset**: use this attribute in cases where dates are not available (e.g., in clinical genomics statements pertaining to a relative of a patient in a family health history)
- **referenceToHealthRecord**: a pointer to an information system (e.g., EHR system) holding the broader contextual data of this observation (this could be implemented by guidance provided by health information exchange systems such as the specifications set out by the NwHIN in the US)

# **Class InterpretivePhenotype (ext. Phenotype)**

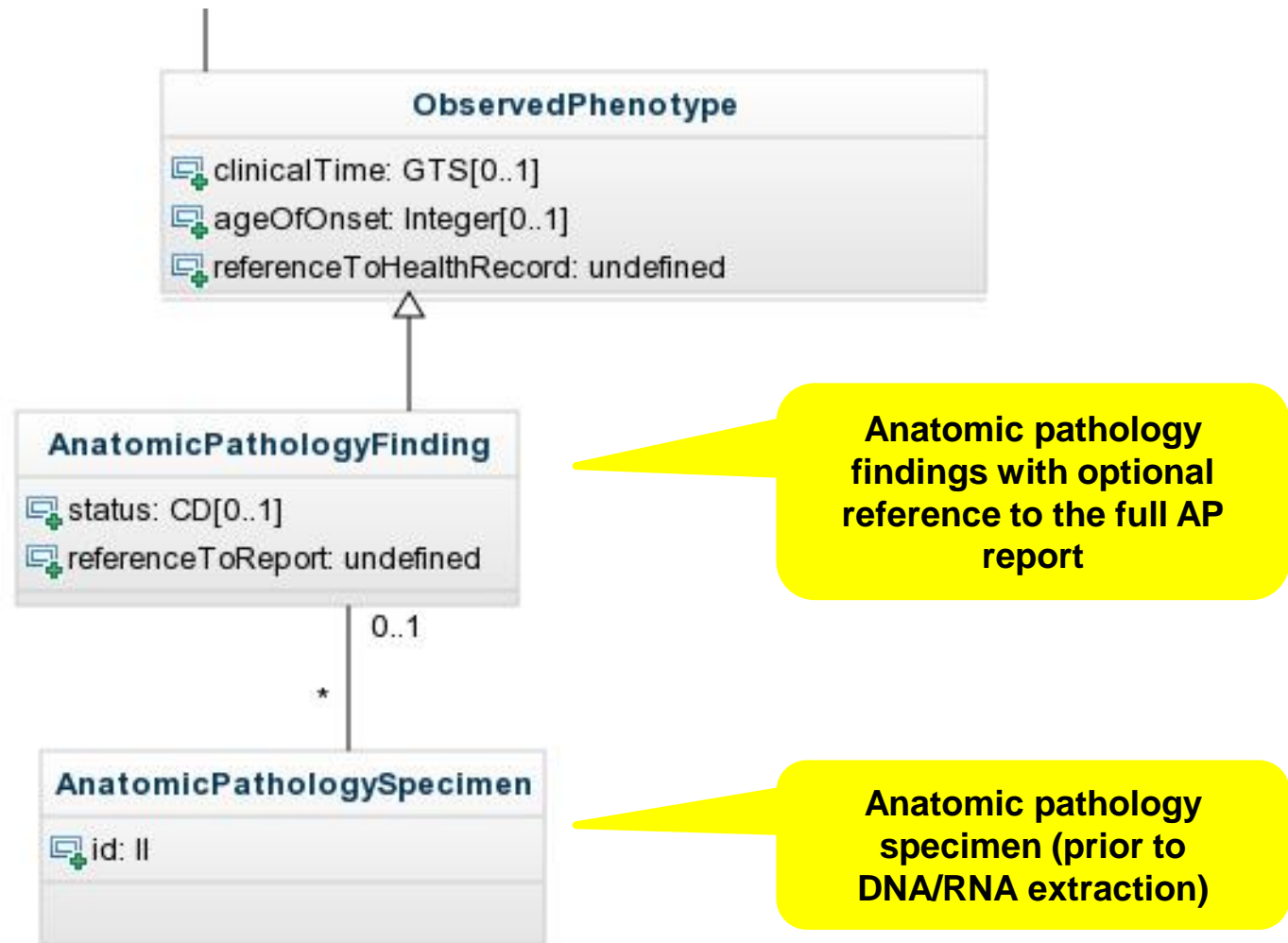
- **Description:**

- This is a sub-class of Phenotype, representing a phenotype that is likely to be manifested in the subject, e.g, the somatic mutation is likely to cause resistance to some drug

- **Attributes:**

- **method:** the method by this interpretation was made (e.g., an identified algorithm, a rule engine, a clinical decision support application/service, etc.)
- **interpretation Time:** the time this interpretation was made available for this Clinical Genomics statement instance)
- **performer:** unique id of the entity making this interpretation (e.g., clinical decision support application/service)

# Anatomic Pathology Specimen & Findings



# The End

- Thank you for your attention... 😊
- Questions? Contact Amnon at [amnon.shvo@gmail.com](mailto:amnon.shvo@gmail.com)
- Comments of general interest should be posted to the CG mailing list at [clingenomics@lists.hl7.org](mailto:clingenomics@lists.hl7.org)