



GEN2PHEN

(<http://www.gen2phen.org>)

Anthony Brookes
University of Leicester, UK

VISION:

**An internet 'Knowledge-Environment'
for G2P information, supporting**

- direct data submission**
- seamless data integration**
- holistic searching**

**FP7 Integrated Project, EUR 12M
January 2008 - December 2012**

**HEALTH-2007-A-1.1.1.0-1:
1.1. HIGH-THROUGHPUT RESEARCH**

**UNIFYING HUMAN AND MODEL
ORGANISM GENETIC VARIATION DATABASES**

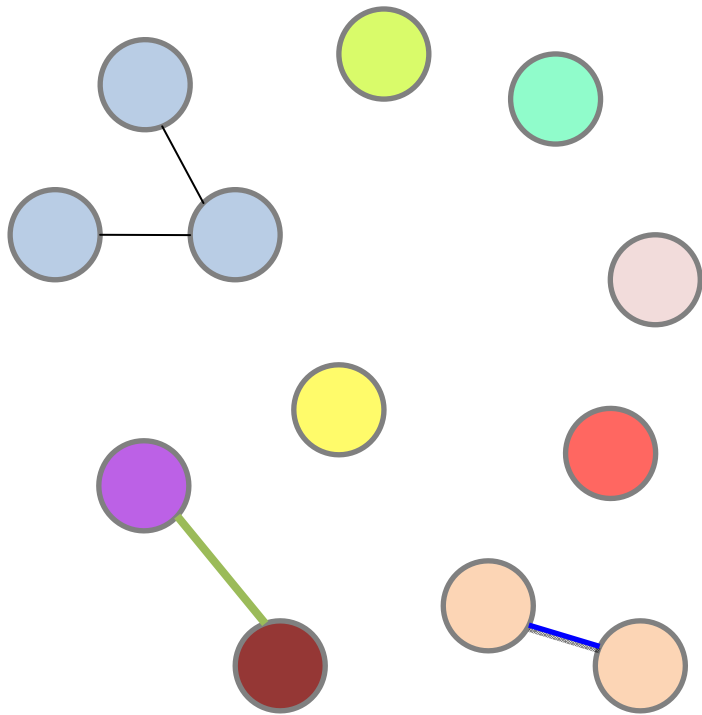
The focus should be on **developing a data and analysis structure** by creating a **hierarchy of bioinformatics grid-linked databases, tools and standards, centred on a generalised existing or novel genome browser**. This major European initiative should **integrate and link databases that contain genotype to phenotype relationships** with the aim of **creating a large integrated genetic variation catalogue**. This should facilitate functional genomics research for human health and where relevant **should take into account international perspectives in the field**.

GEN2PHEN Partners

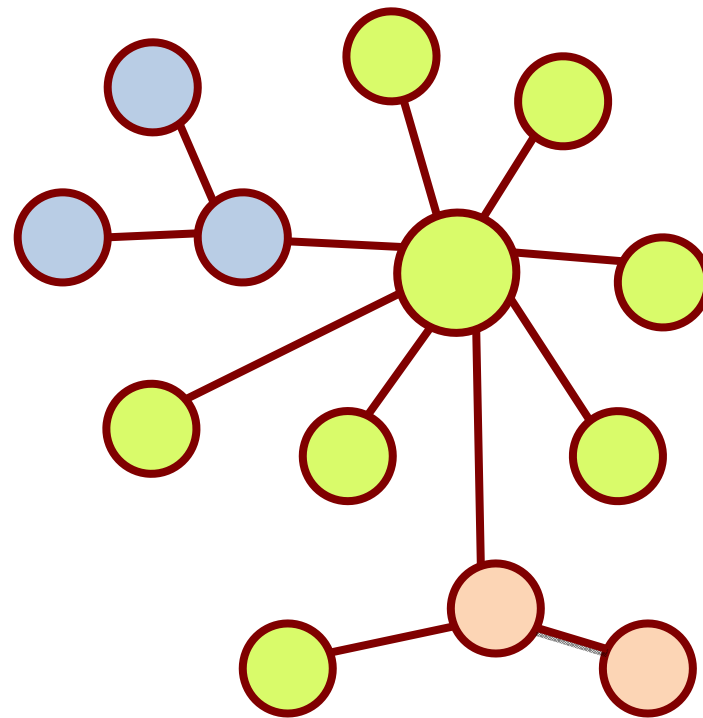
A.J.Brookes	University of Leicester	UK
E.Birney, P. Flicek	European Molecular Biology Laboratory	Germany
C.Díaz	Fundació IMIM	Spain
J.denDunnen	Leiden University Medical Center	Netherlands
C.Béroud	Inst Natl de la Santé et de la Recherche Méd	France
A.Cambon-Thomsen	Inst Natl de la Santé et de la Recherche Méd	France
J-E.Litton	Karolinska Institute	Sweden
G.Potamias	Foundation for Research & Technology	Greece
S.Heath	Centre National de Génotypage	France
G.Patrinou	Erasmus University Medical Center	Netherlands
J.Muilu	University of Helsinki	Finland
J.L.Oliveira	University of Aveiro – IEETA	Portugal
H.Lehvaslaiho	University of Western Cape	South Africa
D.Dash	Institute of Genomics and Integrative Biology	India
L.Yip	Swiss Institute of Bioinformatics	Switzerland
A.Devereau	University of Manchester	UK
A.Kel	BioBase GmbH	Germany
H.Gudbjartsson	deCODE genetics	Iceland
D.Atlan	PhenoSystems	Belgium
T.Kanninen	Biocomputing Platforms	Finland

GEN2PHEN: Core Objectives

- 1: Analyse current needs and practices (global perspective)
- 2: Develop key standards for the G2P field [e.g., PaGE-OM]
- 3: Create generic components, services and integration structures
- 4: Create search and presentation solutions, anchored on Ensembl
- 5: Promote and facilitate data population into G2P databases
- 6: Assist deployment of GEN2PHEN solutions to the community
- 7: Consider system durability and long-term financing
- 8: Execute system validation pilots, with biomedical relevance
- 9: Build a 'G2P nexus' or 'Knowledge Center' (www.gen2phen.org)
(encouraging community input, discussion, and collaboration)



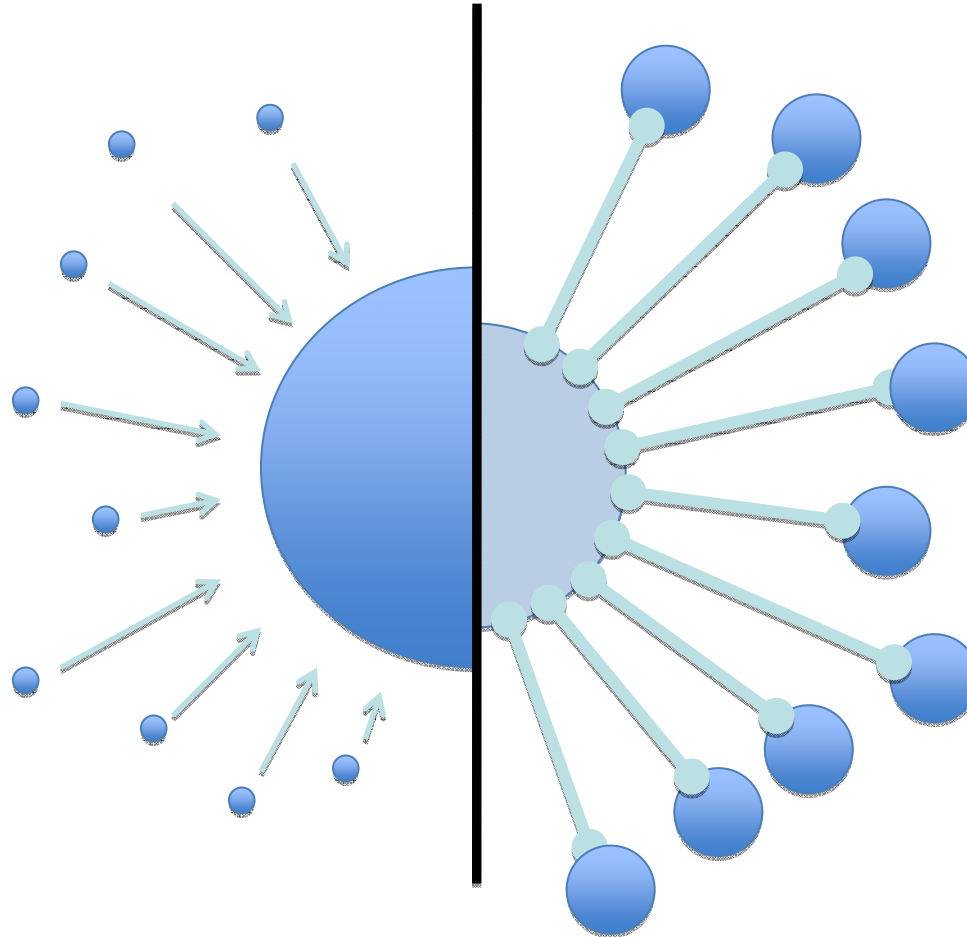
**Current DBs
= disparate silos**



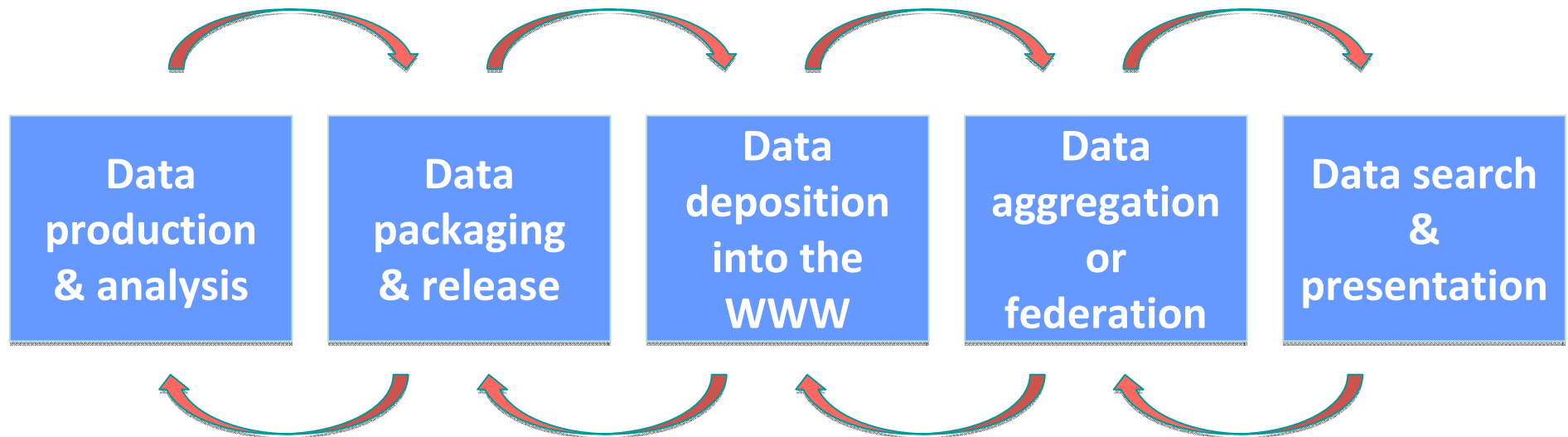
**Future DBs
= federated network**

Central DBs

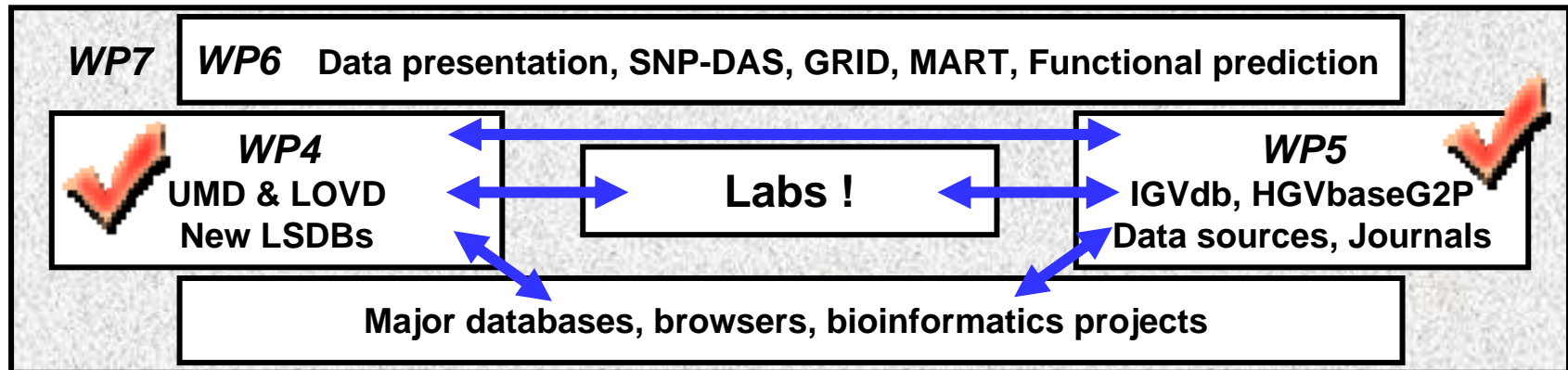
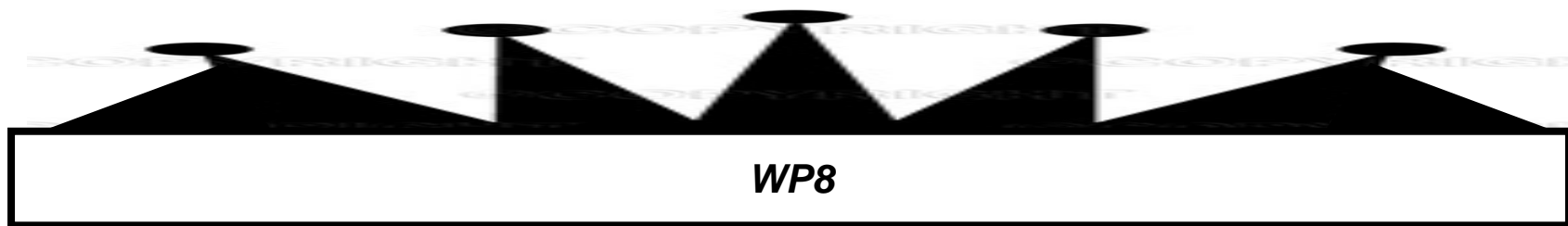
Federated DBs



Data flow



Accreditation and Reward



GEN2PHEN generated standard #1: Locus Reference Genomic Sequences (LRG)

Background

A meeting sponsored by GEN2PHEN (www.gen2phen.org) was held at the European Bioinformatics Institute (EBI), Hinxton, UK on 24–25 April 2008 to discuss a specification for reference genomic DNA sequences that will enable the consistent and unambiguous reporting of mutations.

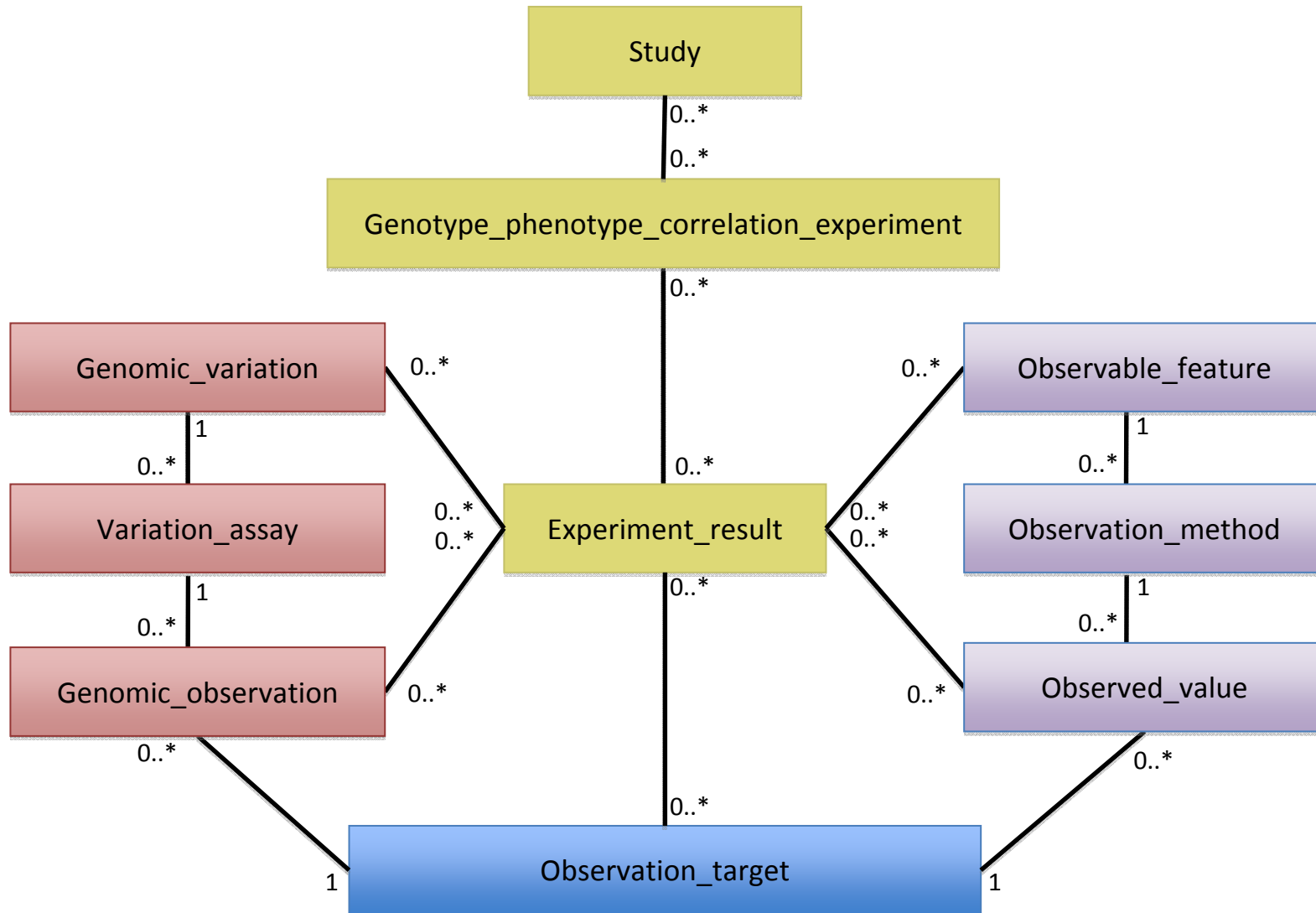
Principles

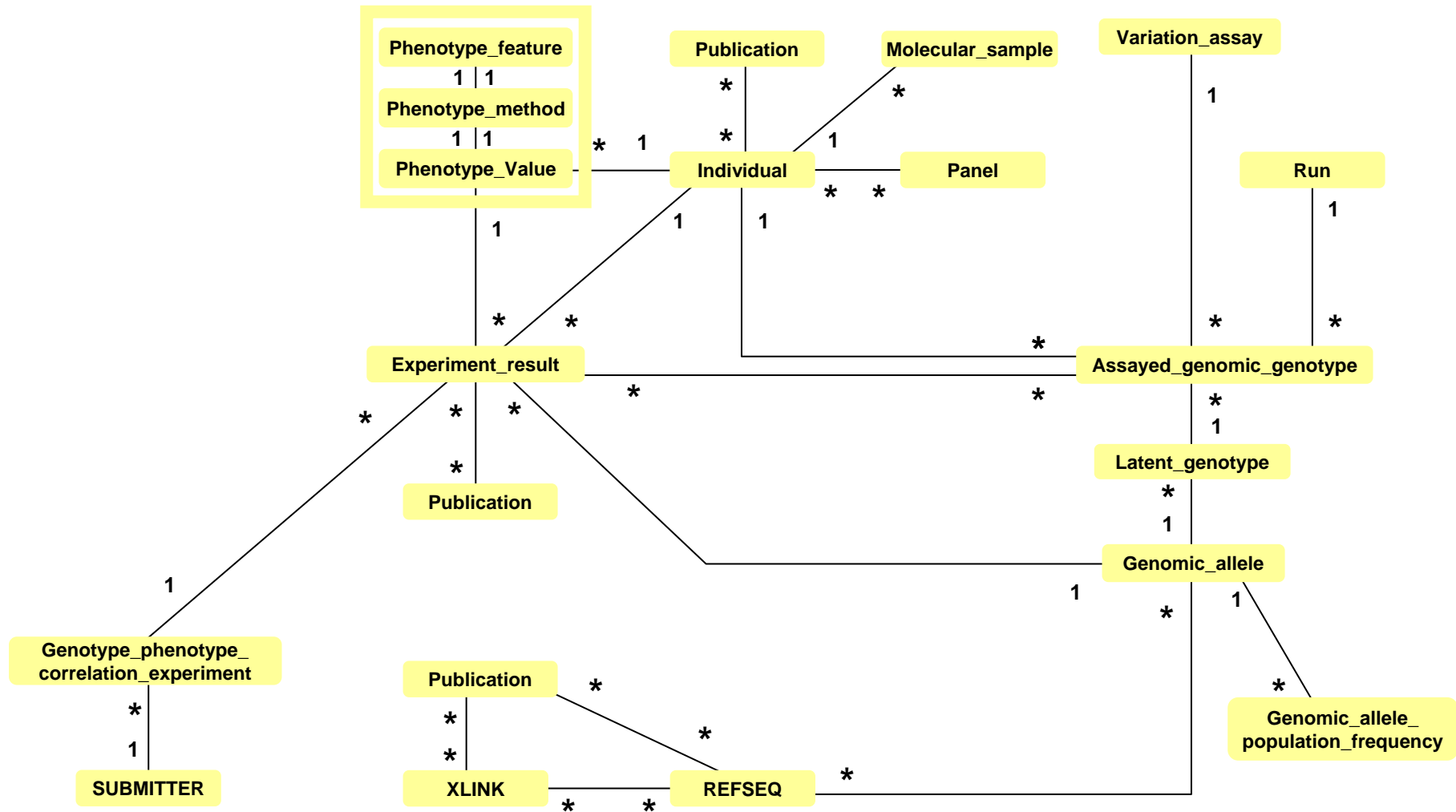
- The community will have the final say in defining the sequences and their annotation
- The sequences need not represent real alleles of genes
- Stability of LRGs is paramount to ensure utility over time frames of many decades

The Solution

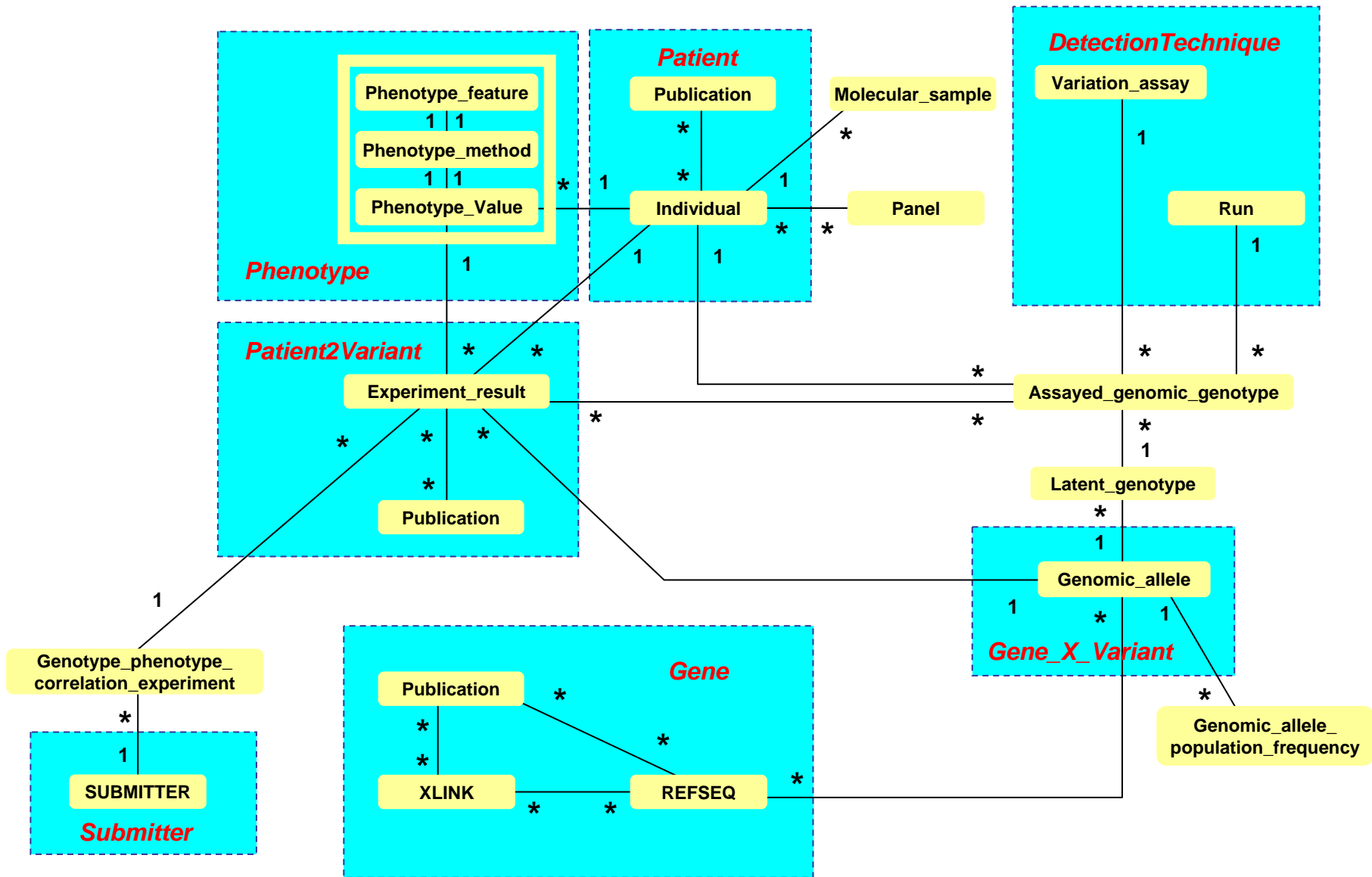
- A genomic DNA sequence representation of a gene, that has a permanent ID (no versioning), and core content that never changes (i.e. nucleotide sequence, exon numbering, exon base positions, start and stop positions).
- LRG is not a nomenclature system (that is provided by the HGVS mutation nomenclature)
- Additional annotations will be present that may change with time, so that the latest ancillary knowledge about a gene is directly available.

**GEN2PHEN generated standard #2:
Phenotype and Genotype Experiment, Object Model (PAGE-OM)
<http://www.pageom.org/>**

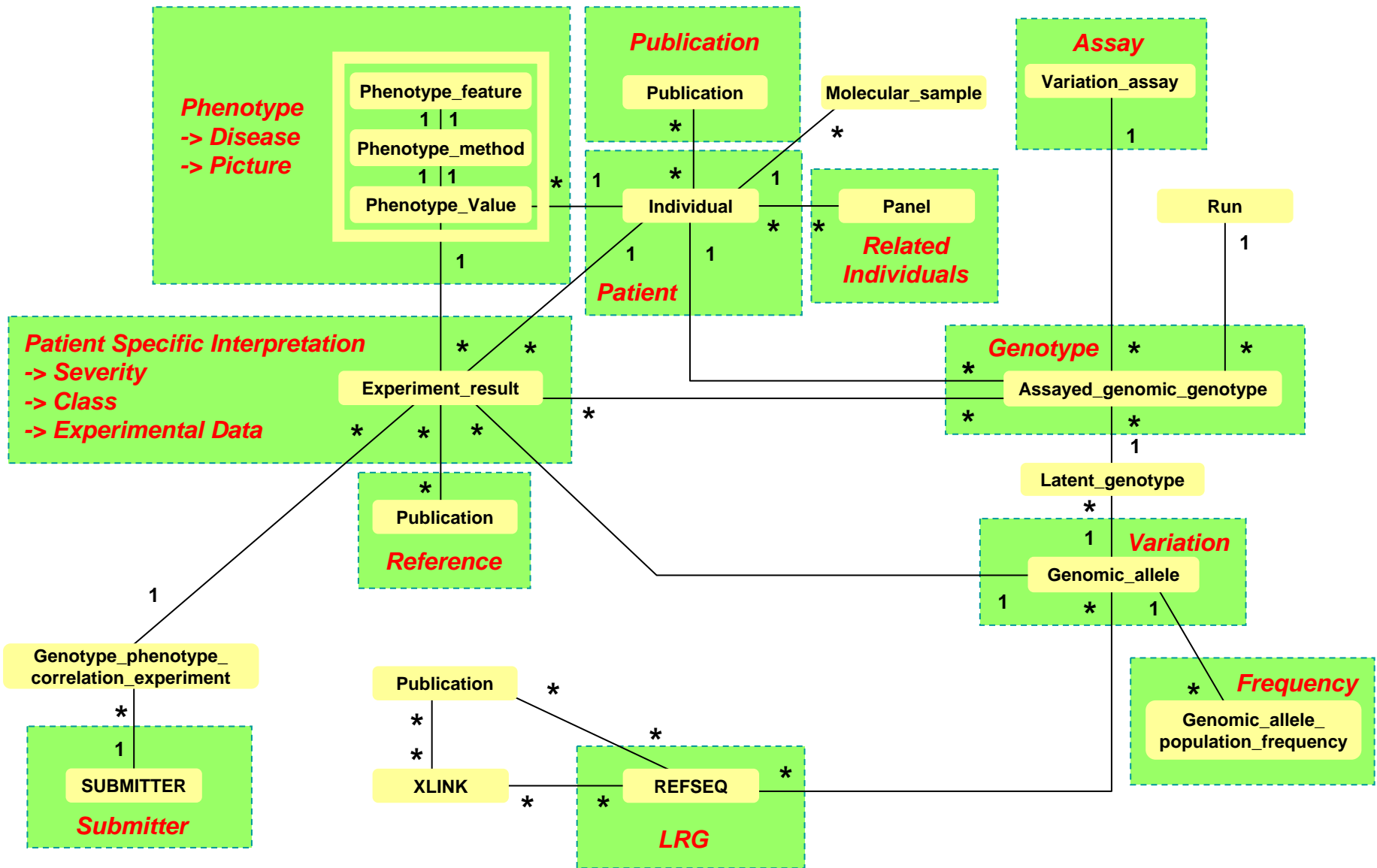




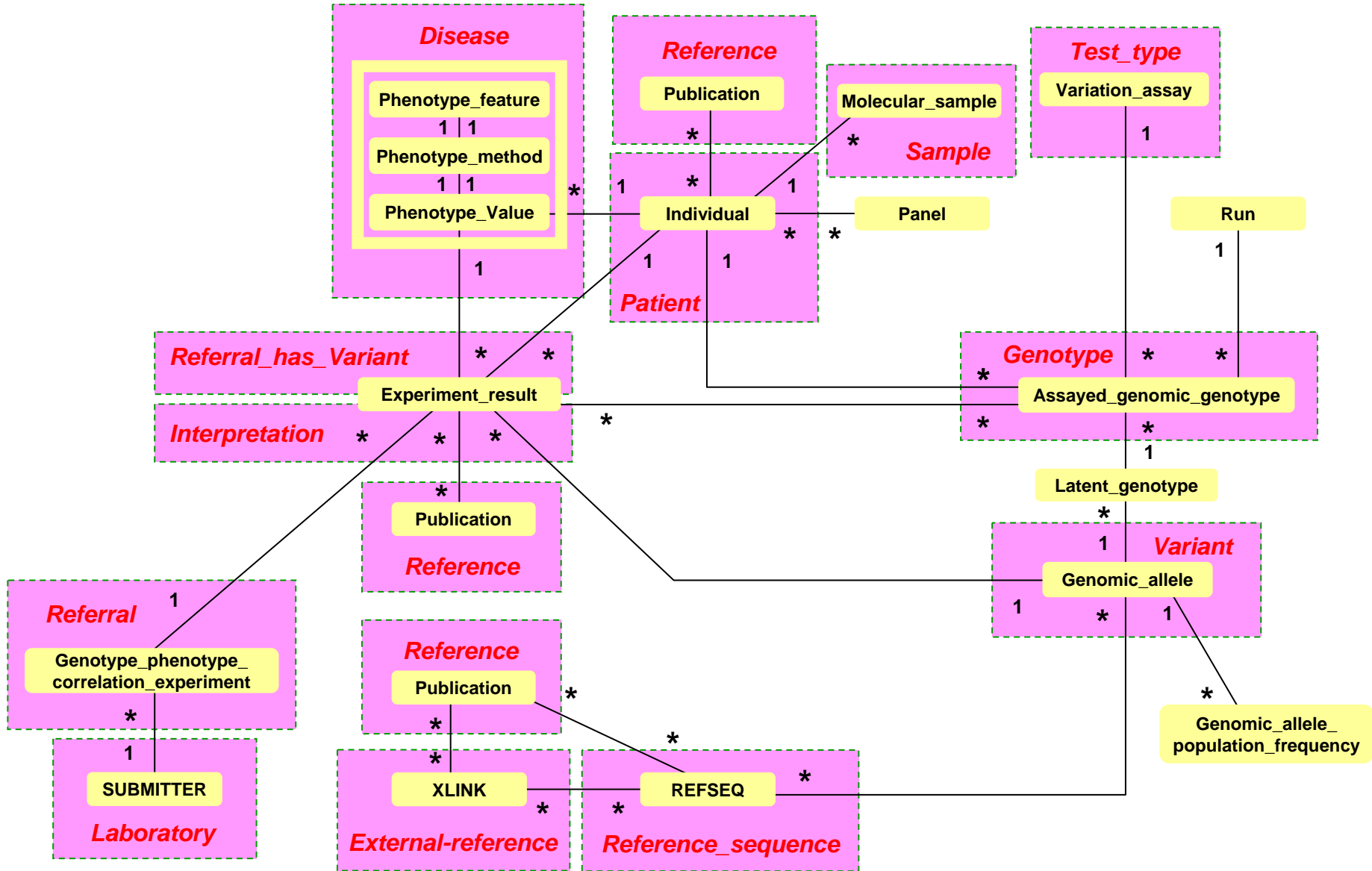
Universal, Core Data Model for LSDBs?



LOVD



UMD



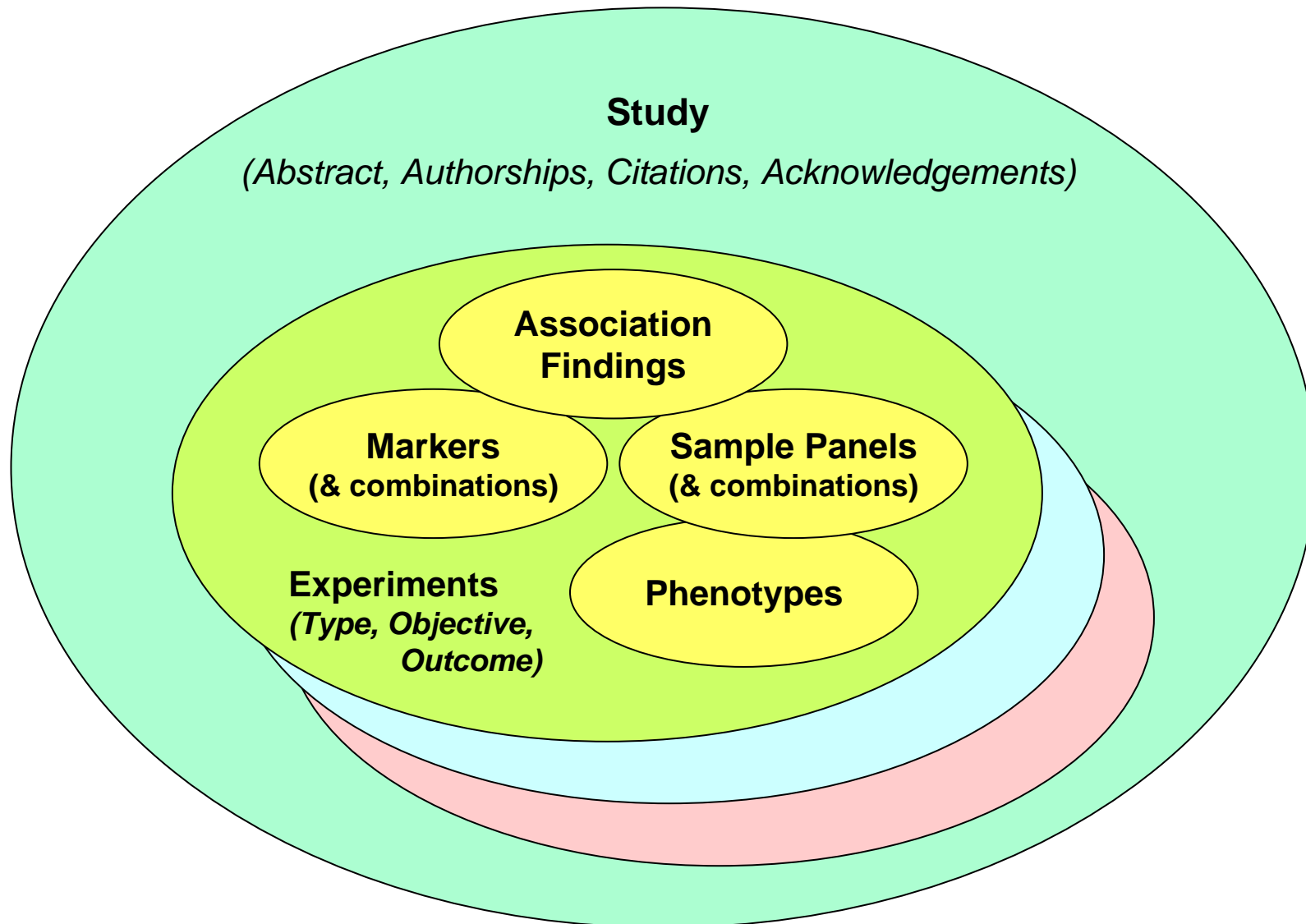
DMuDB

HGVbaseG2P: Human Genome Variation Database, Genotypes To Phenotypes

- **Central genetic association database**
- **Summary level information (no individual data)**
- **Basal layer of all known variants of all types**
- **Active incorporation of large datasets (dbGaP, caBIG, EGA)**
- **Assist submission of datasets from individual labs**
- **Data access via simple search, BioMart, GRID, Ensembl**
- **Data presentation via browse, full record, and graphical views (plus extensive download options)**

MAIN CHALLENGE IS THE INNATE COMPLEXITY OF THE DATA !

...so we conceptualise entries as 'publications', made up of 6 manageable chunks
...and focus database design, searches, and data presentation on these elements



HGVbaseG2P

[Help](#) | [Bulk downloads](#) | [Contact](#) | [Search](#)



[Home](#)

[Studies](#)

[Phenotypes](#)

[Markers](#)

[HGVMart](#)

[About](#) | [Data](#) | [Submission](#) | [Advanced Search](#)

The objective of HGVbase (the Human Genome Variation Database) is to provide an accurate, high utility and ultimately fully comprehensive catalog of normal human gene and genome variation, useful as a research tool to help define the genetic component of human phenotypic variation ([read more](#)).

To get information of specific interest to you, either click on one of the main menu items above or use the quick access panels below:

Search all of HGVbaseG2P

Search across all categories by entering a keyword such as study ID, disease name or author name

Search Studies

[Show me an example](#)
Get information on a study of interest by entering its HGVbaseG2P identifier or keywords.

Or browse the [list of available studies](#).

Latest studies added:
[CGEMS prostate cancer Test study](#)

Search Phenotypes

[Show me an example](#)
Retrieve data related to phenotype properties (or traits) and the methods used to measure or ascertain them in GWA studies.

Or browse the [list of available phenotypes](#).

Search Markers

[Show me an example](#)
Get marker information by entering an HGVbaseG2P or external marker identifier.

Or search the database with our [advanced marker search](#).

Get data

Use HGVMart to obtain...[rephrase]

Choose focus:

Or download summary-level genotype data and associations in bulk from our [Data page](#).

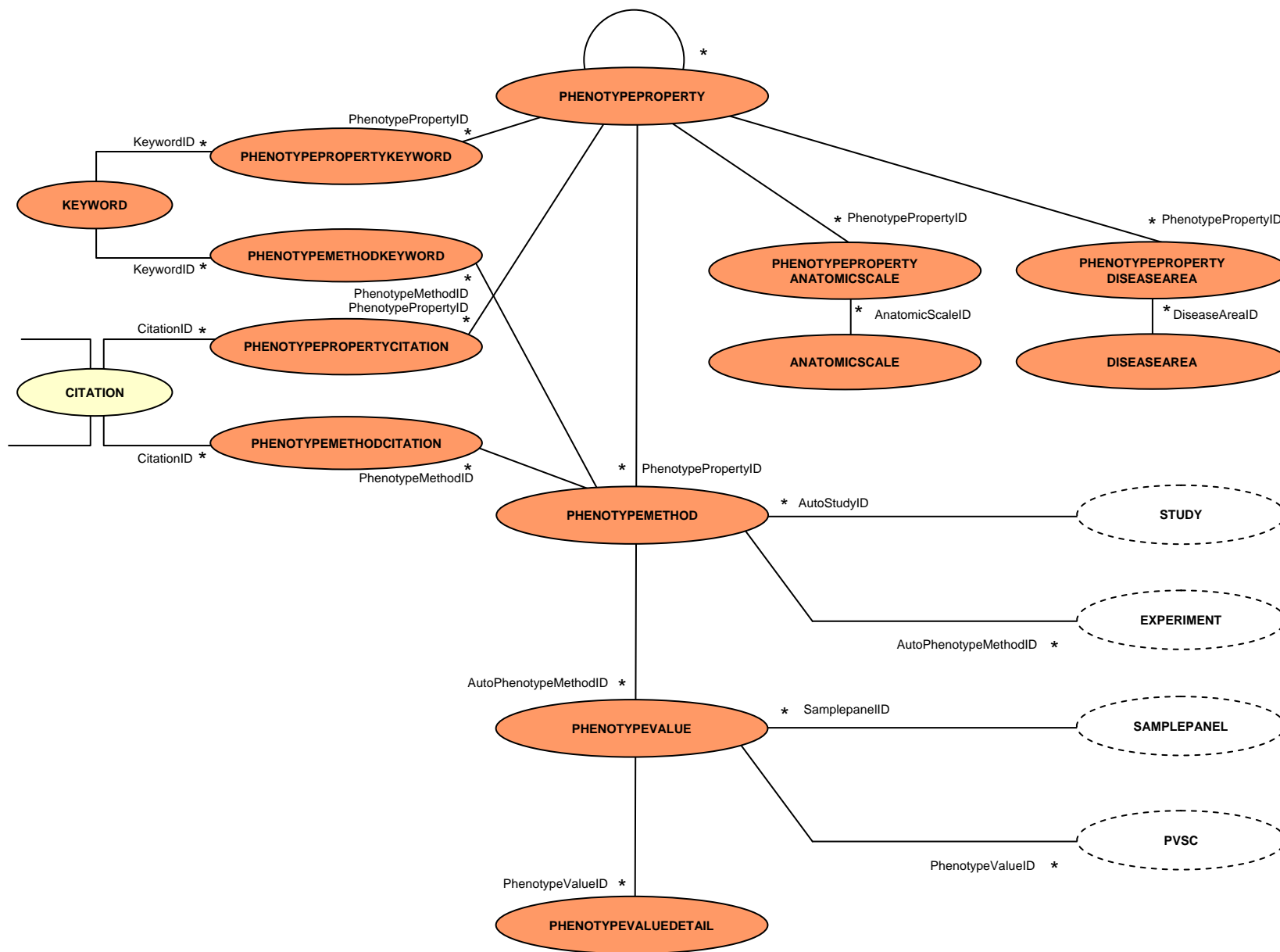
Launch Content (July 2008)

Summary level genotype and association data:

- **dbSNP b.129 :** Markers and Allele frequency data (>19 million Markers)
- **Broad DGI:** Diabetes Genetics (3,000 Scandinavians, 386,731 Markers)
- **Finland/USA:** FUSION NIDDM (1161 cases, 1174 controls, ~315,000 Markers)
- **NIH NINDS:** Parkinson's Disease (267 PD cases, 270 controls, ~400,000 Markers)
- **NIH NEI:** Age-Related Eye Disease Study (350 cases, 171 controls, ~107,000 Markers)
- **NCI CGEMS:** Breast Cancer (1145 cases, 1142 controls, ~ 550,000 markers)
Prostate Cancer & Follow-up (1177 cases, 1105 controls, ~300,000 markers)
- **Welcome Trust:** Case Control Consortium (14,000 cases, 3,000 controls, ~500,000 markers)
Tuberculosis
Coronary heart disease
Type 1 diabetes
Type 2 diabetes
Rheumatoid arthritis
Crohn's disease
Bipolar disorder
Hypertension

Post-launch Content

- **Constantly:** **Semi-automated gathering from major depositories**
- **2009 onwards:** **Any & all ‘complete’ datasets, of any size, submitted by the community**
- **2010 onwards:** **GRID access to ‘HGVBaseG2P-in-a-box’ installations globally (e.g., ENGAGE)**



UTILITY



GEN 2 PHEN

Acknowledgements

GEN2PHEN Consortium

21 groups, 19 Institutions

Brookes Lab

Gudmundur Thorisson, Owen Lancaster, Johan Klinberg

Marianne Siegfried, Rob Free, Rob Hastings

Collaborators

Samir Brahmachari, Debasis Dash, et al (IGIB, India)

Heikki Lehtväslaiho (SANBI, South Africa)

PaGE-OM Consortium, esp.

Kimitoshi Naito et al (Japan)

Hideaki Sugawara et al (Japan)

Juha Muilu (Finland)

Martin Senger (Philippines)

Funding:

EC FP6 NoE INFOBIOMED

EC FP7 IP GEN2PHEN

GlaxoSmithKline, UK