

GEN2PHEN Partners (www.gen2phen.org)

Academic

A.J.Brookes, R.Dalgleish	University of Leicester	UK
P.Flicek, H.Parkinson	European Molecular Biology Laboratory	Germany
C.Díaz	Fundació IMIM	Spain
J.denDunnen	Leiden University Medical Centre	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.Potam		
G.Patrin		
S.Heath		
J.Muilu		
J.L.Olive		
D.Dash		
L.Yip		
A.Dever		

**...towards an internet
'Knowledge-Environment' for
G2P information**

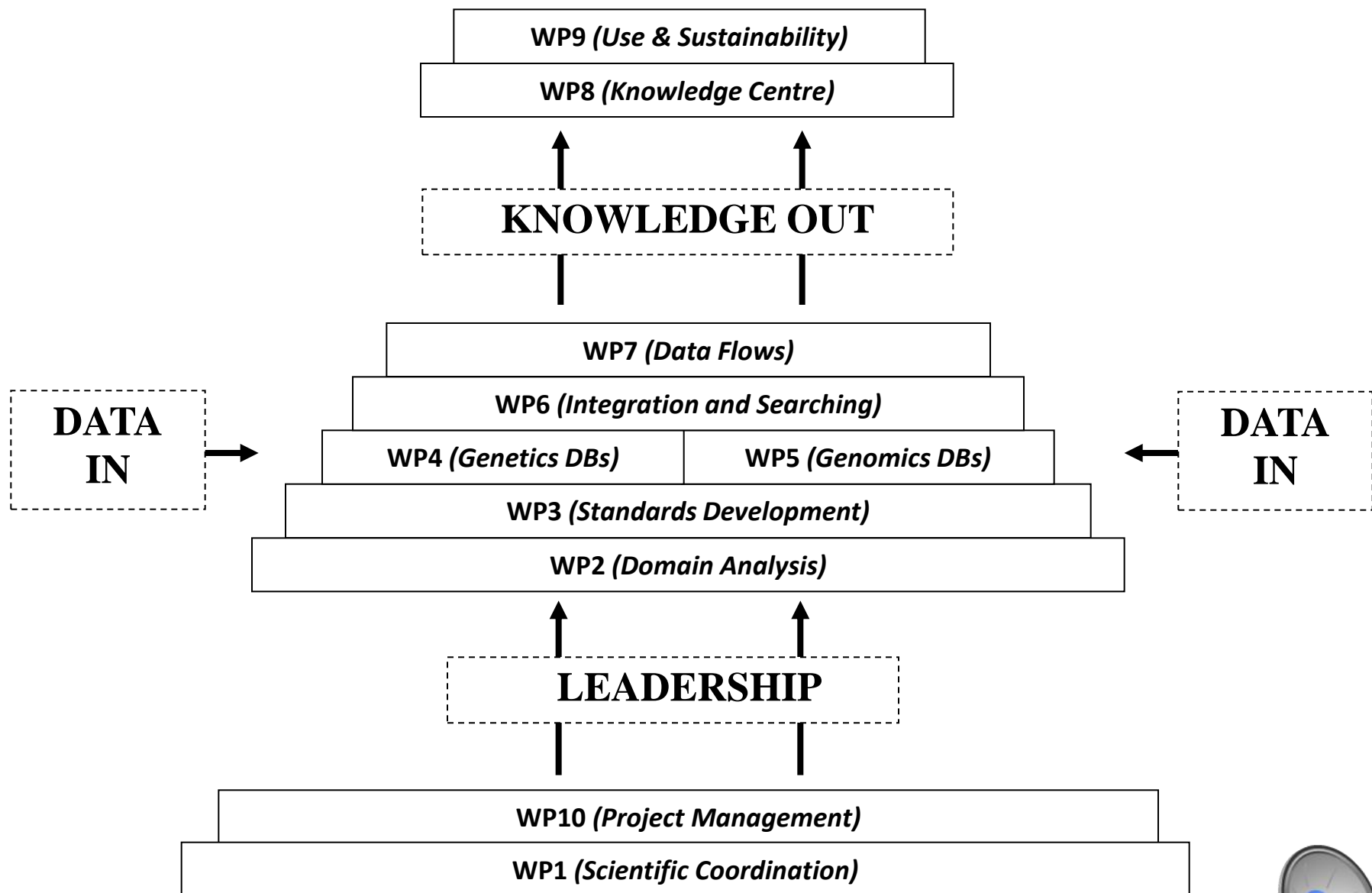
SMEs

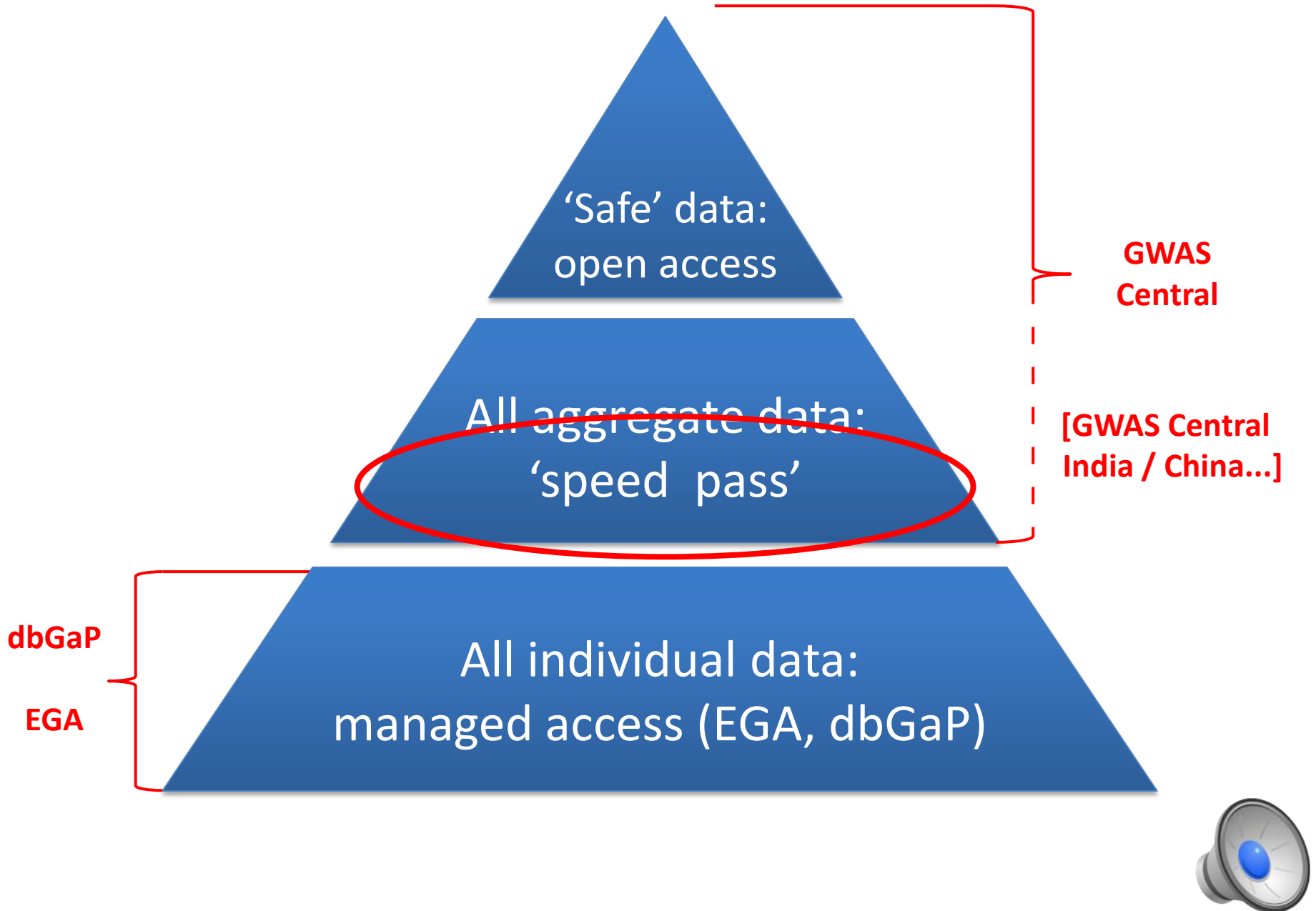
A.Kel	BioBase GmbH	Germany
H.Gudbjartsson	deCODE genetics	Iceland
D.Atlan	PhenoSystems	Belgium
T.Kanninen	Biocomputing Platforms	Finland

Associates

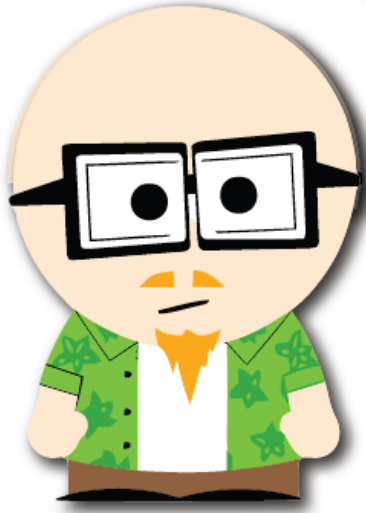
H.Lehvaslaiho	University of Western Cape	South Africa
M.Swertz	Groningen University Medical Centre	Netherlands
M.Vihinen	University of Tampere	Finland







RESEARCHER IDENTIFIERS:



ORCID ID: B-1242-2010

G. Thorisson, Univ. Leicester

G. A. Thorisson, Univ. Leicester

G. A. Thorisson, Cold Spring Harbor Lab.

unique, permanent, not reused !

...but, you can have more than one !





*Openly share the 'existence' rather than
the 'substance' of the data
....thereafter variably manage data access*



OPEN data sharing: ...more than one way!

Anthony Brookes
University of Leicester, UK



PUBLIC DOMAIN GENOME BROWSERS

e.g. Ensembl

**...a seamless internet
'Knowledge-Environment'
for biomedical information**

DNA + Sequence
databases

Model organism
G2P databases

Locus-specific
databases + Diagnostics
databases

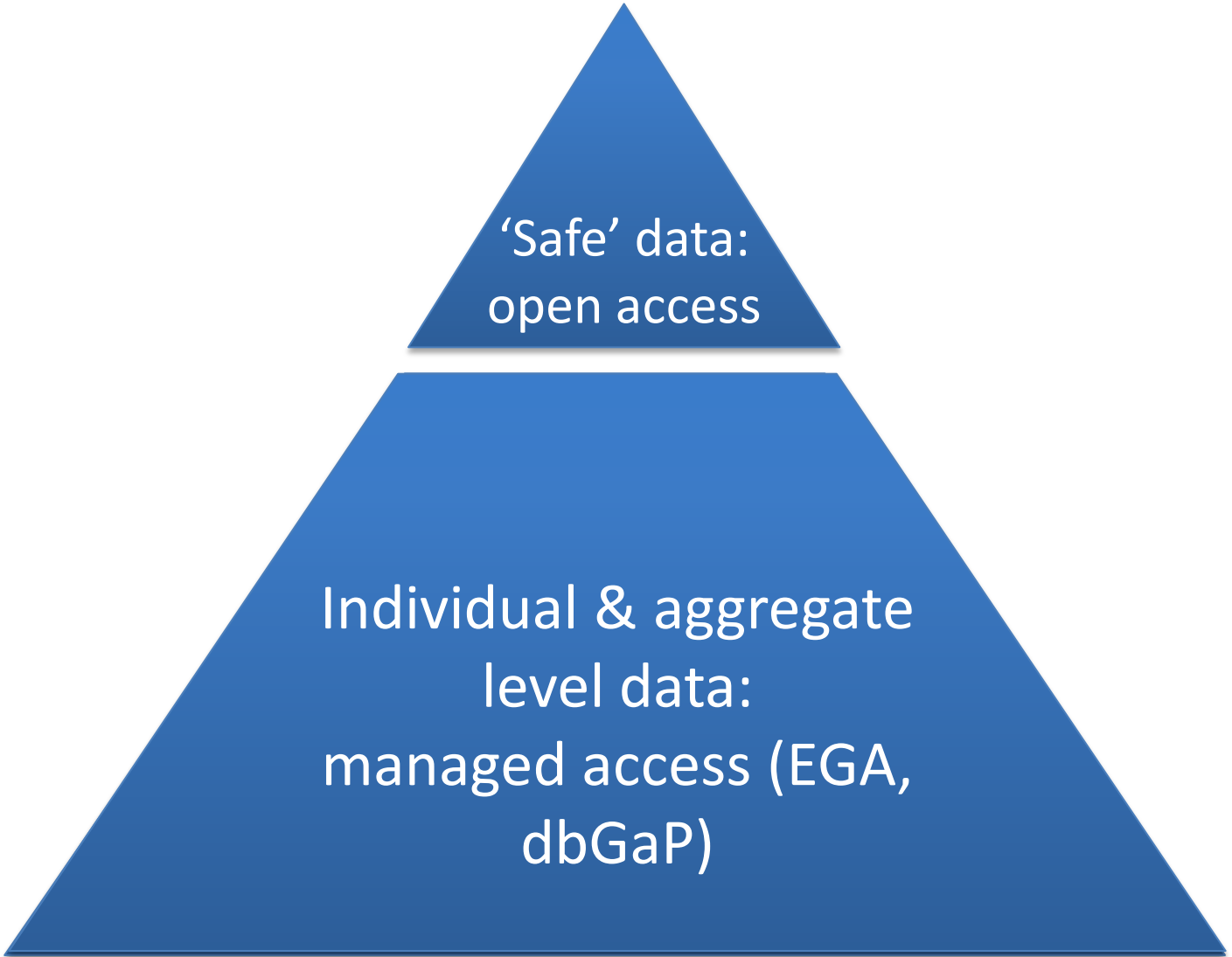
Summary-level
genomics + Individual-level
G2P Databases

GEN2PHEN activities...

- 1: Analyse current needs and practices (global perspective)
- 2: Develop key standards for the G2P field
- 3: Create generic components, services and integration structures
- 4: Create search and presentation solutions, anchored on Ensembl
- 5: Assist deployment of GEN2PHEN solutions, and federate
- 6: Promote and facilitate data population into G2P databases
- 7: Consider system durability and long-term financing

Issues that restrict sharing data

- Researchers may not have time nor funding to manually submit data, and/or submission process and requirements too complicated
- Researchers receive little or no recognition or reward for releasing data, hence little incentive to try
- Researchers may have positive reasons for NOT wanting to share data (ethical, legal, competitive edge)
- No current SANCTIONS for researchers that do not maximally share data



'Safe' data:
open access

Individual & aggregate
level data:
managed access (EGA,
dbGaP)

Enter a study or marker identifier, keywords, a gene name, or chromosomal region.
(e.g. [BRCA1](#), [chr12:13234...4534534](#), [12p13.33](#), [cancer](#), [rs2317951](#))



Studies



Phenotypes



AGCGATC
Markers



Browser



Download



About HGvbaseG2P

The Human Genome Variation Database (HGvbaseG2P) provides summary level findings from large and small. We actively gather data from public projects, and encourage the community. [See more..](#)

- *genetic association database*
- *aiming to integrate many datasets*
- *summary level data only*
- *links to data sources for primary data*

up to [Mailing list](#) or [RSS](#)

base update [Read](#)

website

3..

study database 8 [Read](#)

2010/08/05 | HGvbaseG2P releases study database 7 [Read](#)



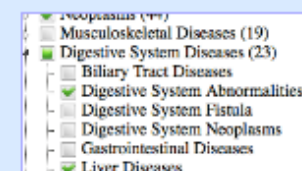
Frequently Asked Questions

- How is the database content organised?
- How do I find Studies of interest?
- How do I find Markers of interest?
- How do I use the Browser to identify regions of interest?
- How do I submit my own data to HGvbaseG2P?

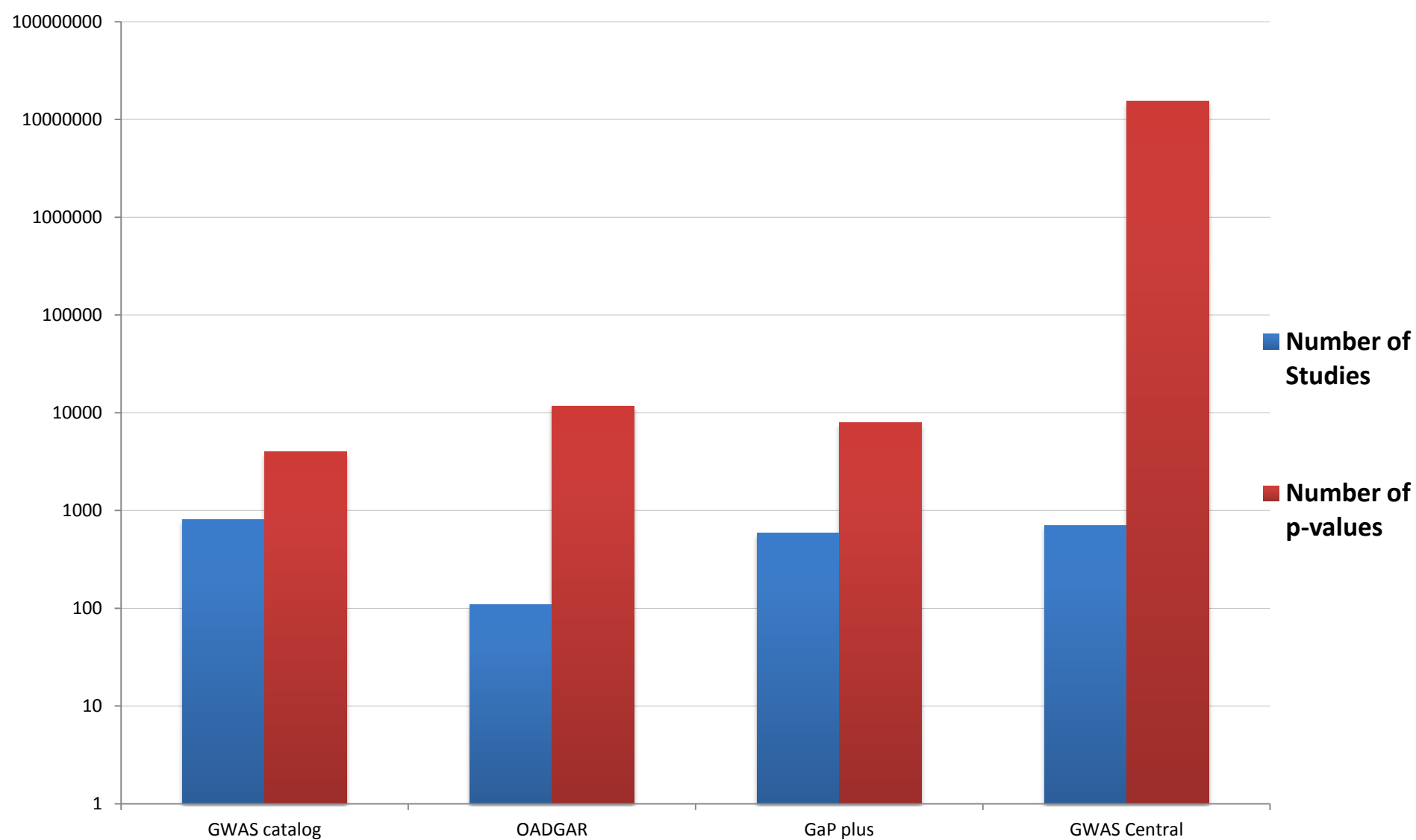


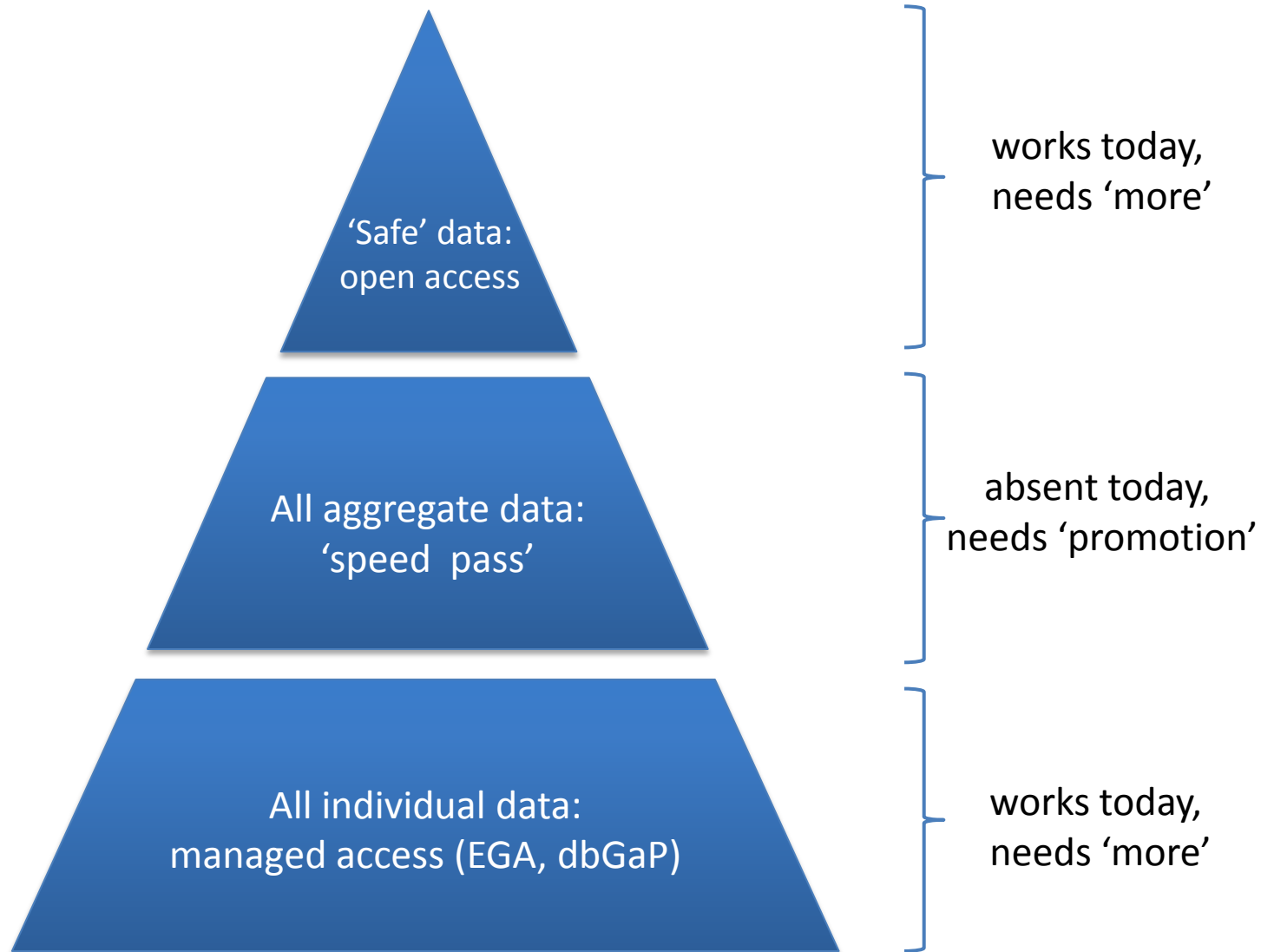
Phenotype Trees

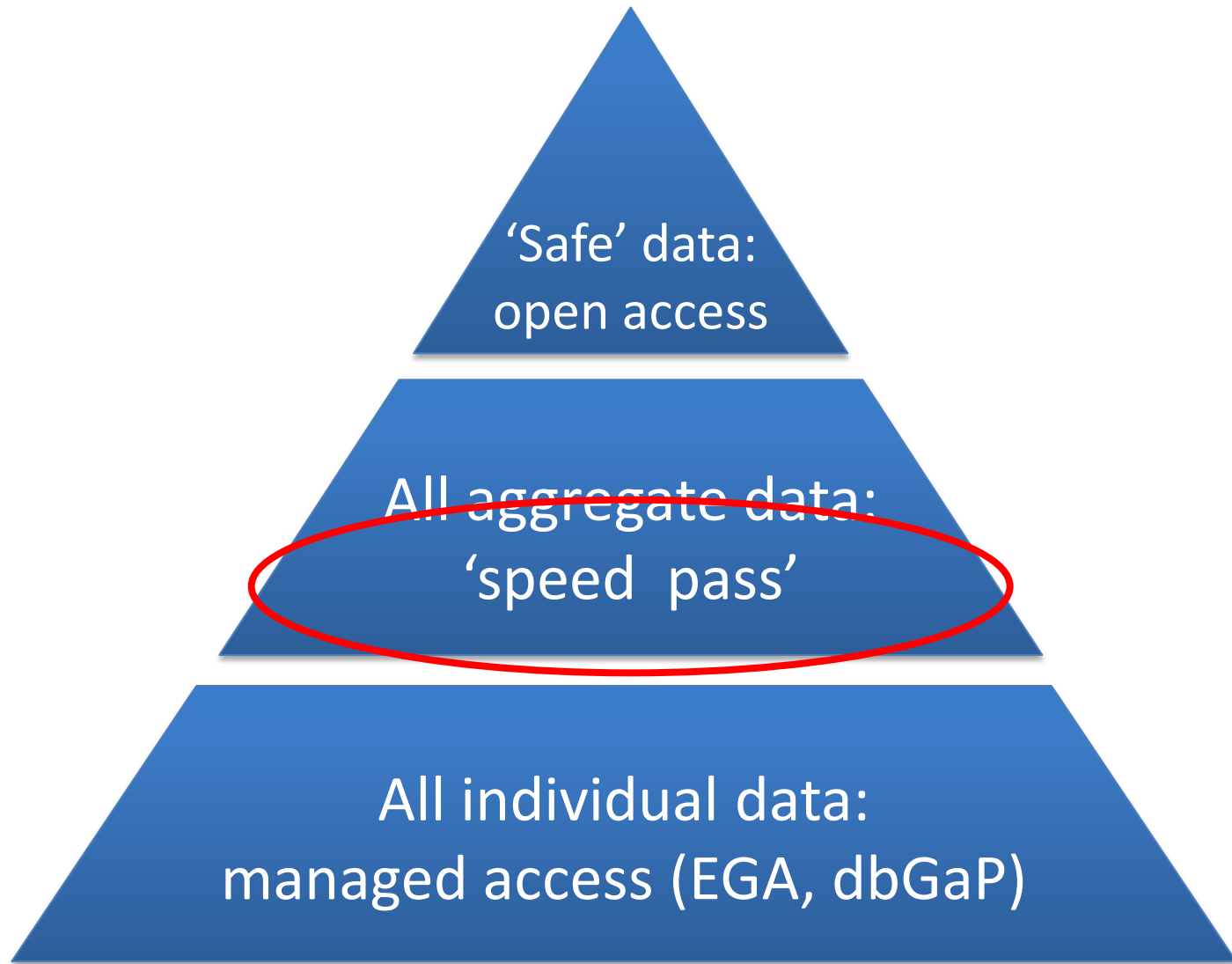
MeSH and HPO based [Phenotype trees](#) allow you to easily find Studies that relate to a specific disease or type of disease.

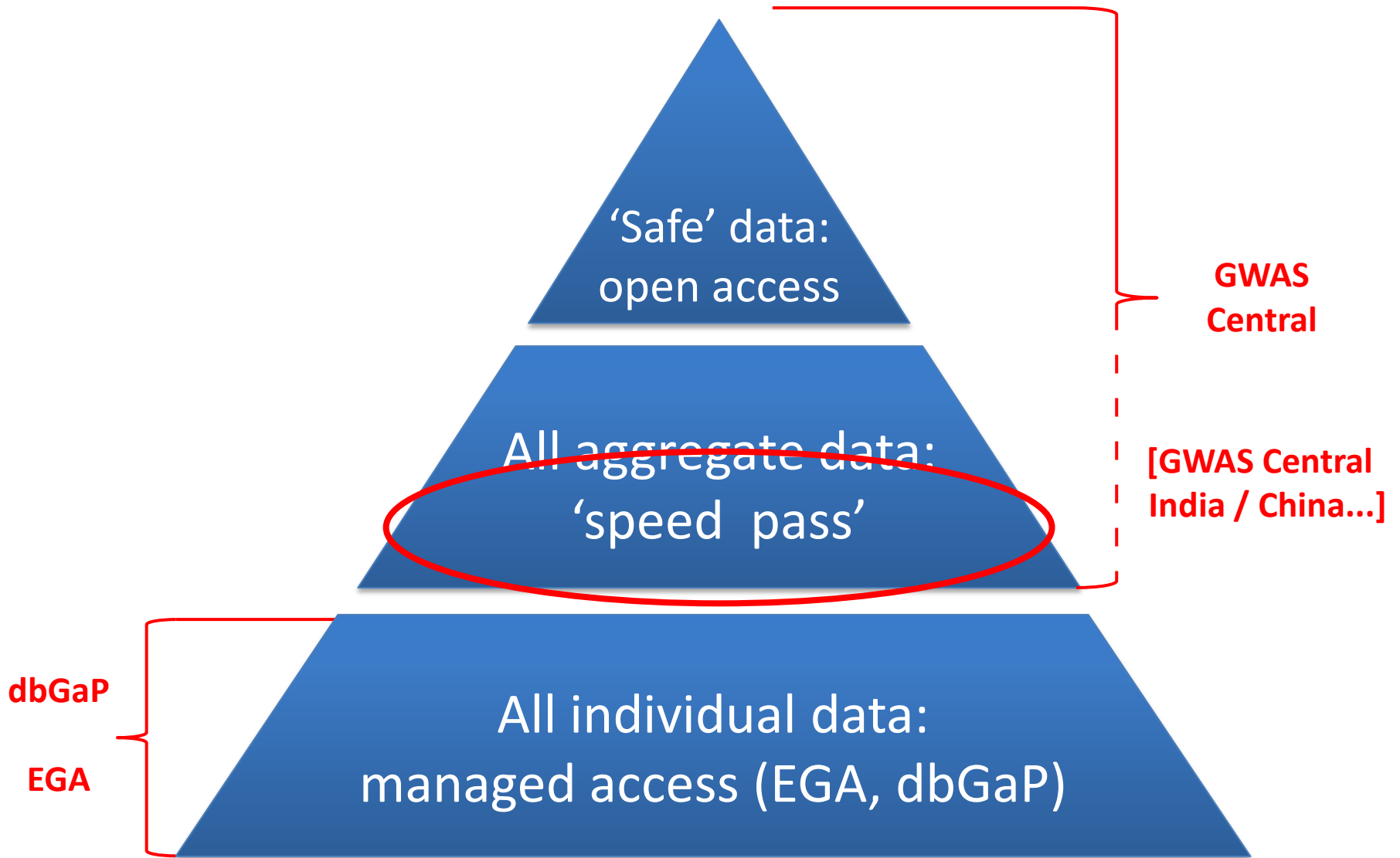


GWAS Central data content compares well with other resources

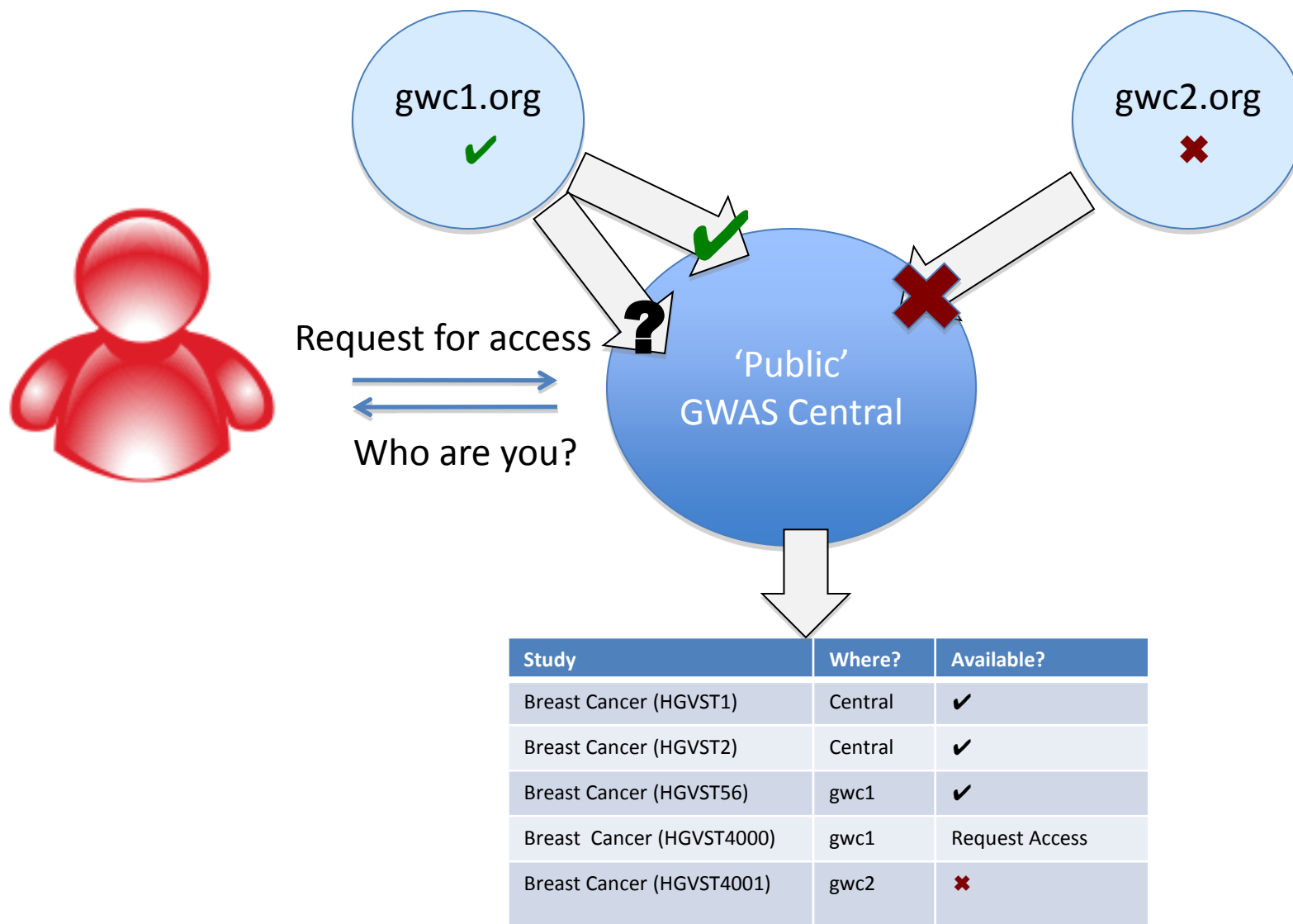


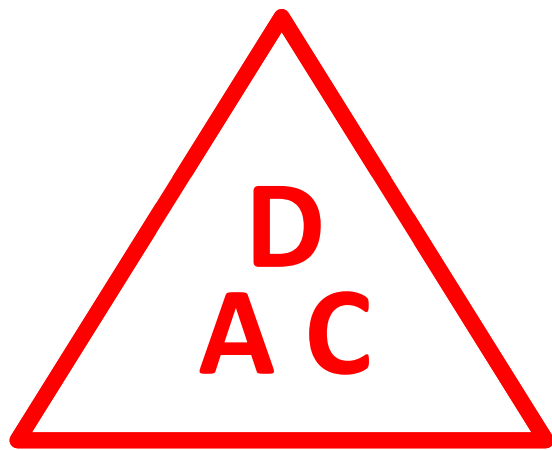






'Federated' GWAS Central



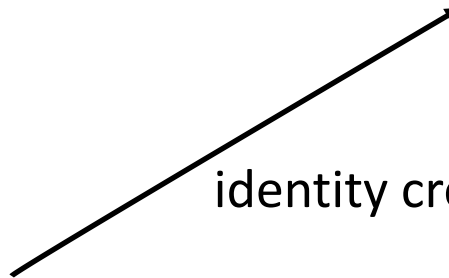
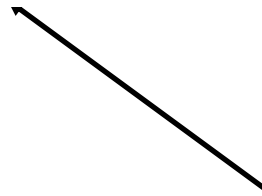


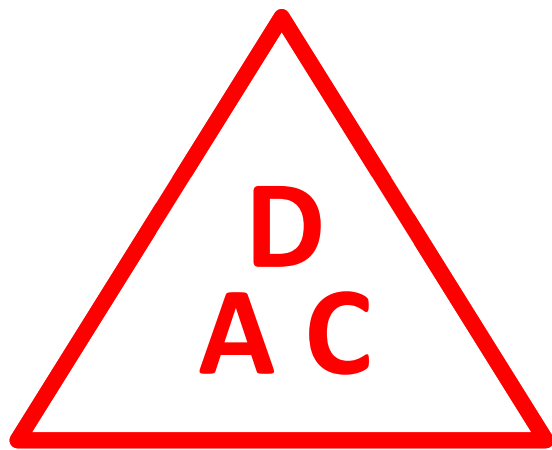
Resource



User

identity credentials



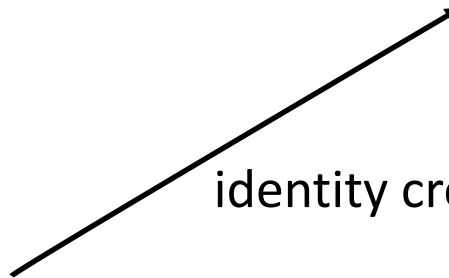
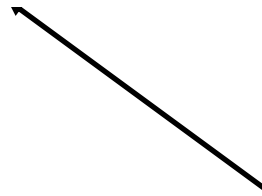


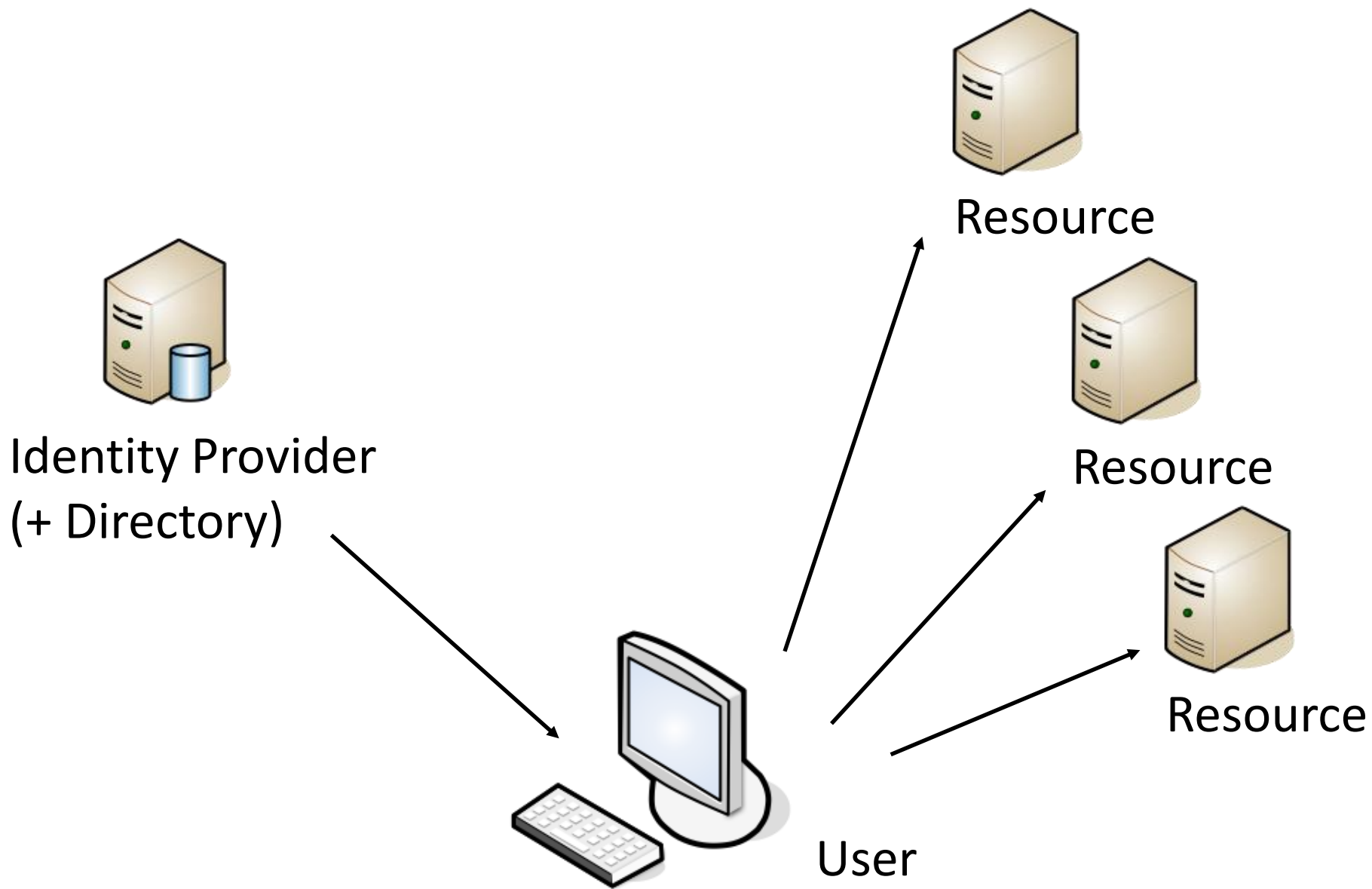
Resource



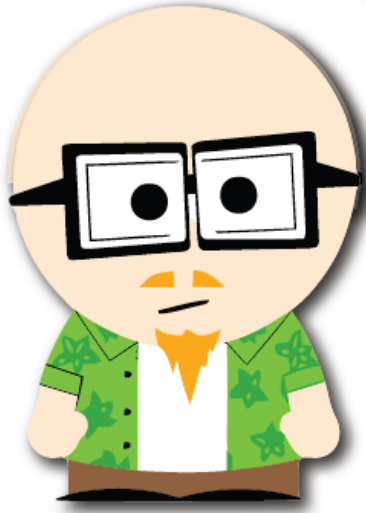
User

identity credentials





RESEARCHER IDENTIFIERS:



ORCID ID: B-1242-2010

G. Thorisson, Univ. Leicester

G. A. Thorisson, Univ. Leicester

G. A. Thorisson, Cold Spring Harbor Lab.

unique, permanent, not reused !

...but, you can have more than one !

Unique identifiers for authors and other contributors

~2/3 of the ~6 million authors in MEDLINE share a last name and first initial with at least one other author, and an ambiguous name refers to ~8 persons on average.

Torvik and Smalheiser. Author name disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data* (2009) vol. 3 (3)

Dec'09: launch of the Open Research Contributor Identification Initiative - ORCID

The screenshot shows the homepage of the Open Research Contributor Identification Initiative (ORCID). The header features the ORCID logo and the text "Working together to align the global network" and "An independent, community effort to standardize researcher identification". Below the header is a navigation bar with links: HOME, ABOUT US, REGISTER, and MEMBER GALLERY. The main content area is divided into several sections. On the left, there is a section titled "THE INITIATIVE" with a blue background image. The text describes the problem of name ambiguity and attribution in the research ecosystem and the mission of the ORCID Initiative to resolve this by assigning unique identifiers. A "MORE" link is provided at the bottom of this section. In the center, there is a section titled "SHARE YOUR THOUGHTS AND PERSONAL STORIES ON NAME AMBIGUITY" with a photo of a person and a molecular model. Below this is a "Coming soon!" section with a downward arrow. On the right, there is a section titled "Participate As Member Organization" with a "REGISTER" button. Below this is a section titled "What about ResearcherID?" with a "ResearcherID" logo and text explaining that many researchers are registered in ResearcherID and that Thomson Reuters will prepare FAQs to help delineate ResearcherID from the new Open Initiative. A "News" section is visible at the bottom left.

Open Research
CONTRIBUTOR IDENTIFICATION INITIATIVE

HOME ABOUT US REGISTER MEMBER GALLERY

THE INITIATIVE

Name ambiguity and attribution are persistent, critical problems imbedded in the scholarly research ecosystem. The ORCID Initiative represents a community effort to establish an open, independent registry that is adopted and embraced as the industry's de facto standard. Our mission is to resolve the systemic name ambiguity, by means of assigning unique identifiers linkable to an individual's research output, to enhance the scientific discovery process and improve the efficiency of funding and collaboration.

[MORE](#)

News

Working together to align the global network
An independent, community effort to standardize researcher identification

SHARE YOUR THOUGHTS AND PERSONAL STORIES ON NAME AMBIGUITY

Coming soon!

Please look for our **Facebook** and **Twitter** presence in early 2010. You will be able to follow the Initiative, provide feedback and suggestions on its future, and hear from peers and colleagues.

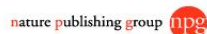
Participate As Member Organization

ORCID: Register your organization's interest in participation [REGISTER](#)

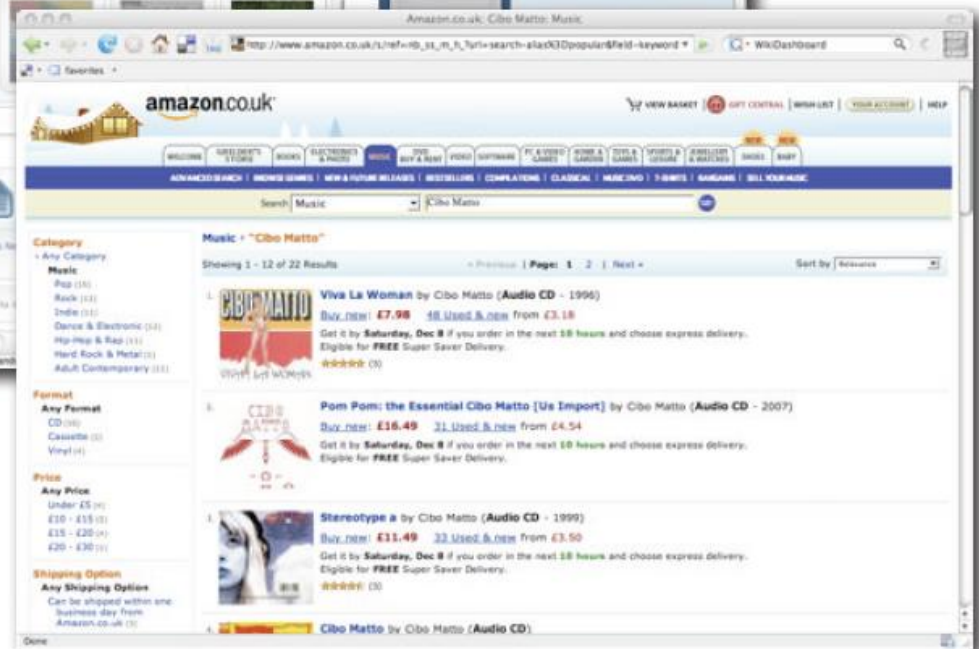
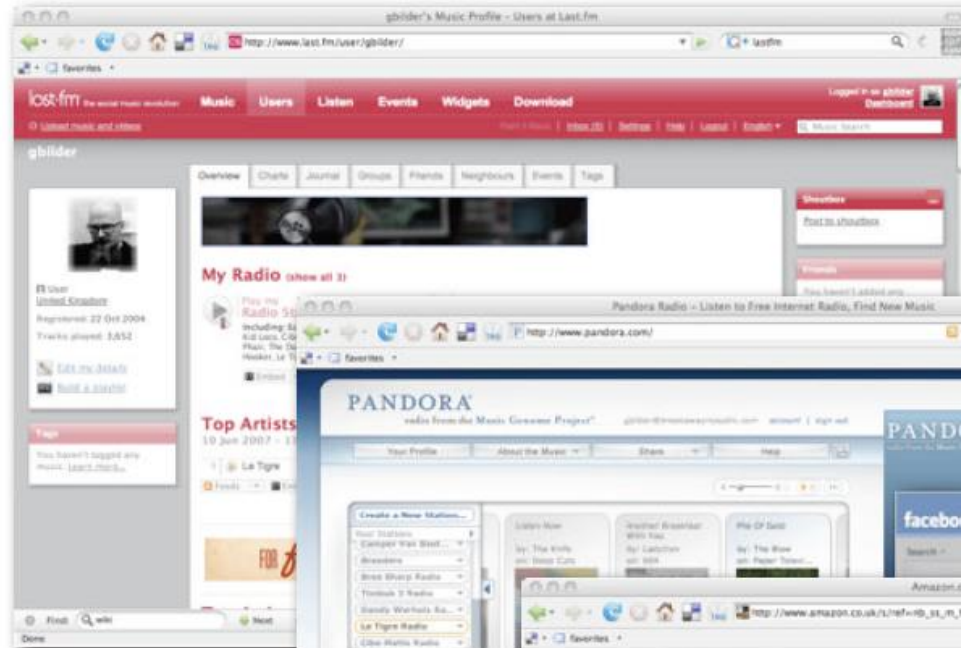
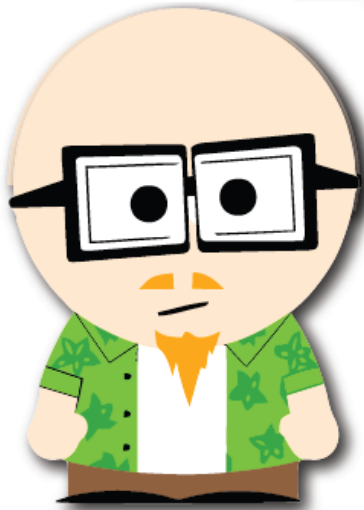
What about ResearcherID?

ResearcherID

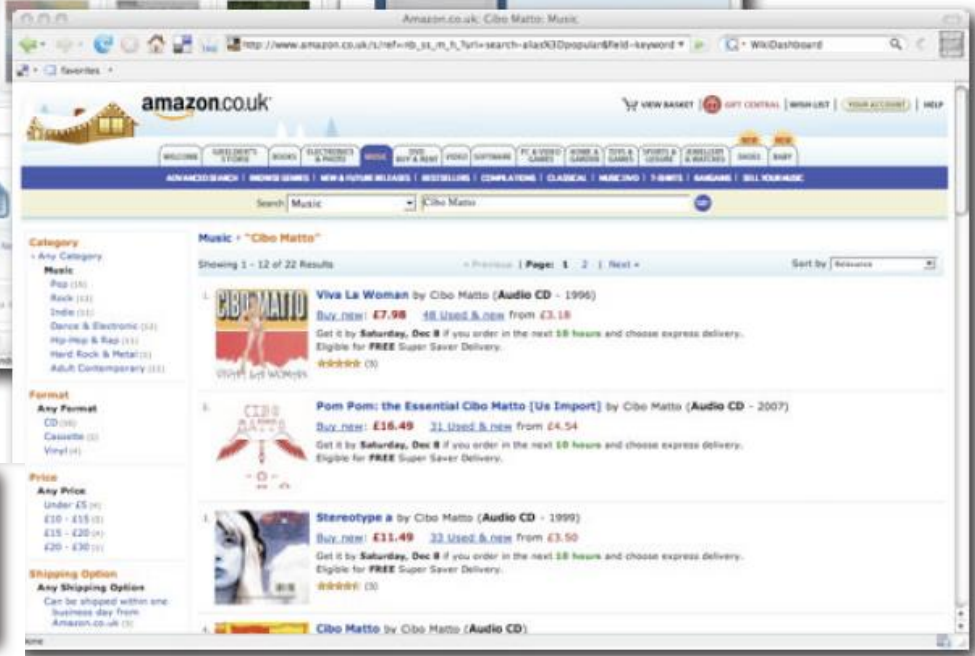
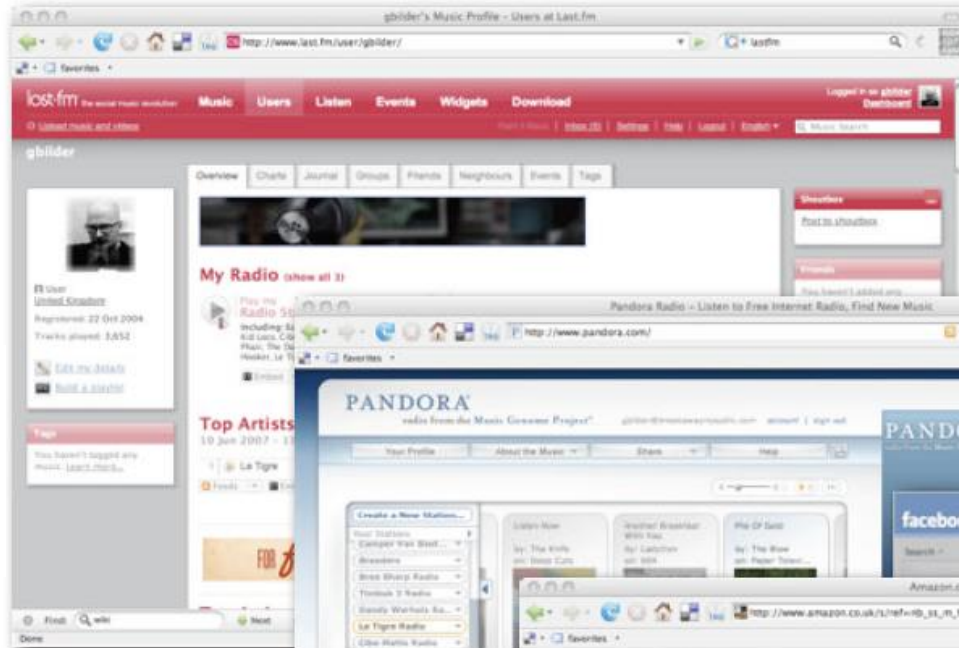
Many researchers are registered in *ResearcherID*. Thomson Reuters will prepare **FAQs** to help delineate *ResearcherID* from the new Open Initiative. Please check back soon.

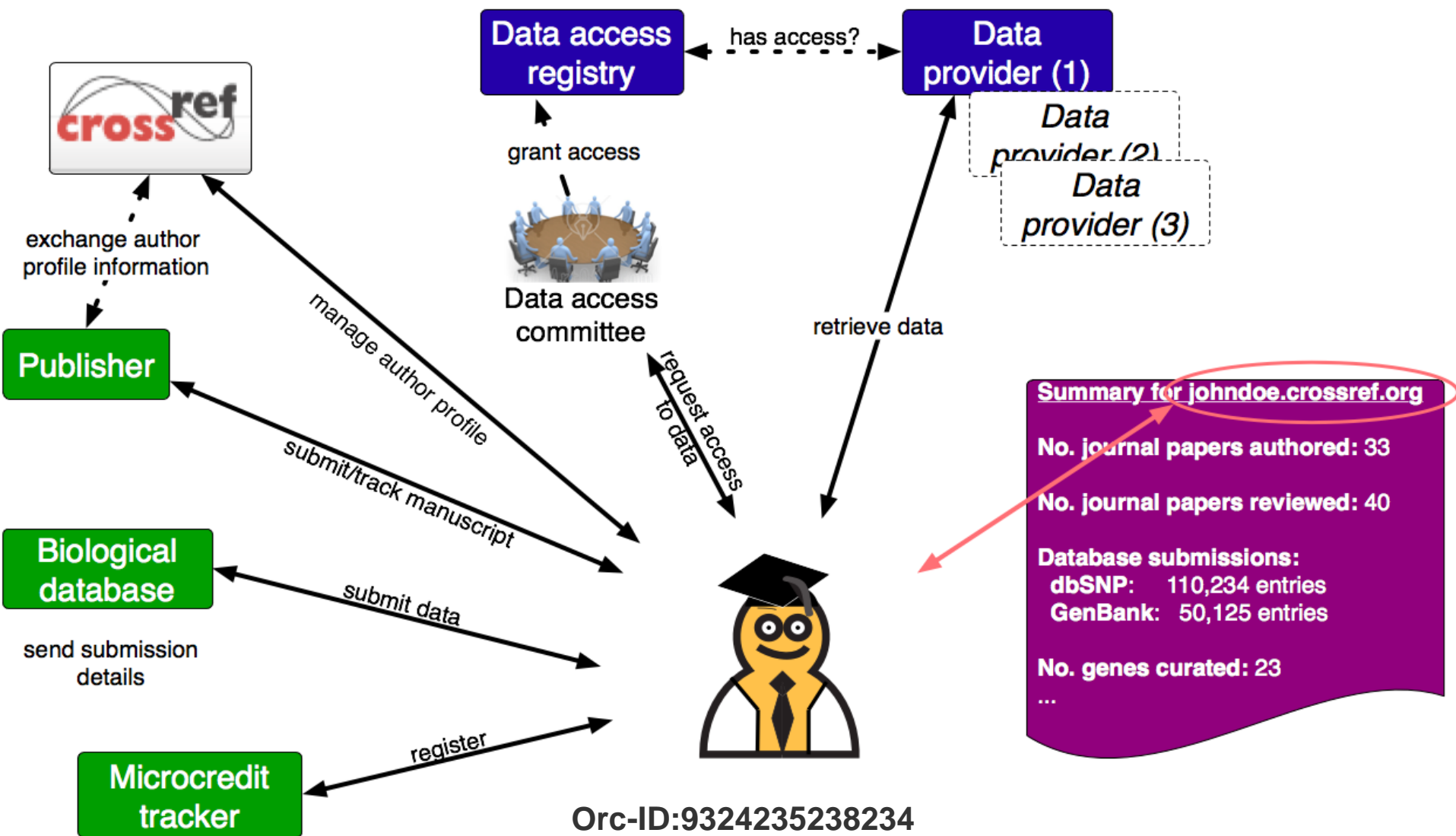


IDENTITY:



IDENTITY:





Orc-ID:9324235238234

G. Thorisson, Univ. Leicester

G. A. Thorisson, Univ. Leicester

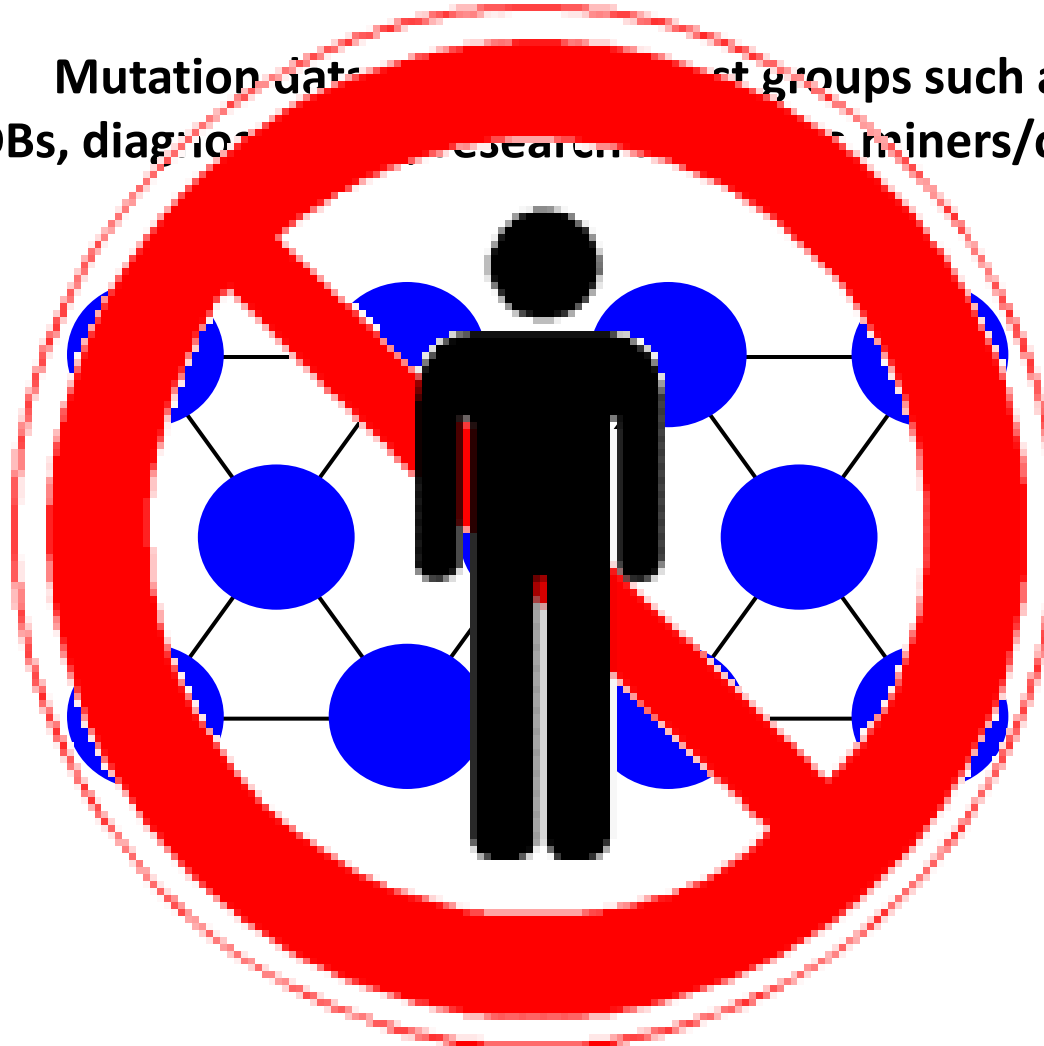
G. A. Thorisson, Cold Spring Harbor Lab.



*Openly share the 'existence' rather than
the 'substance' of the data
....thereafter variably manage data access*

The problem...

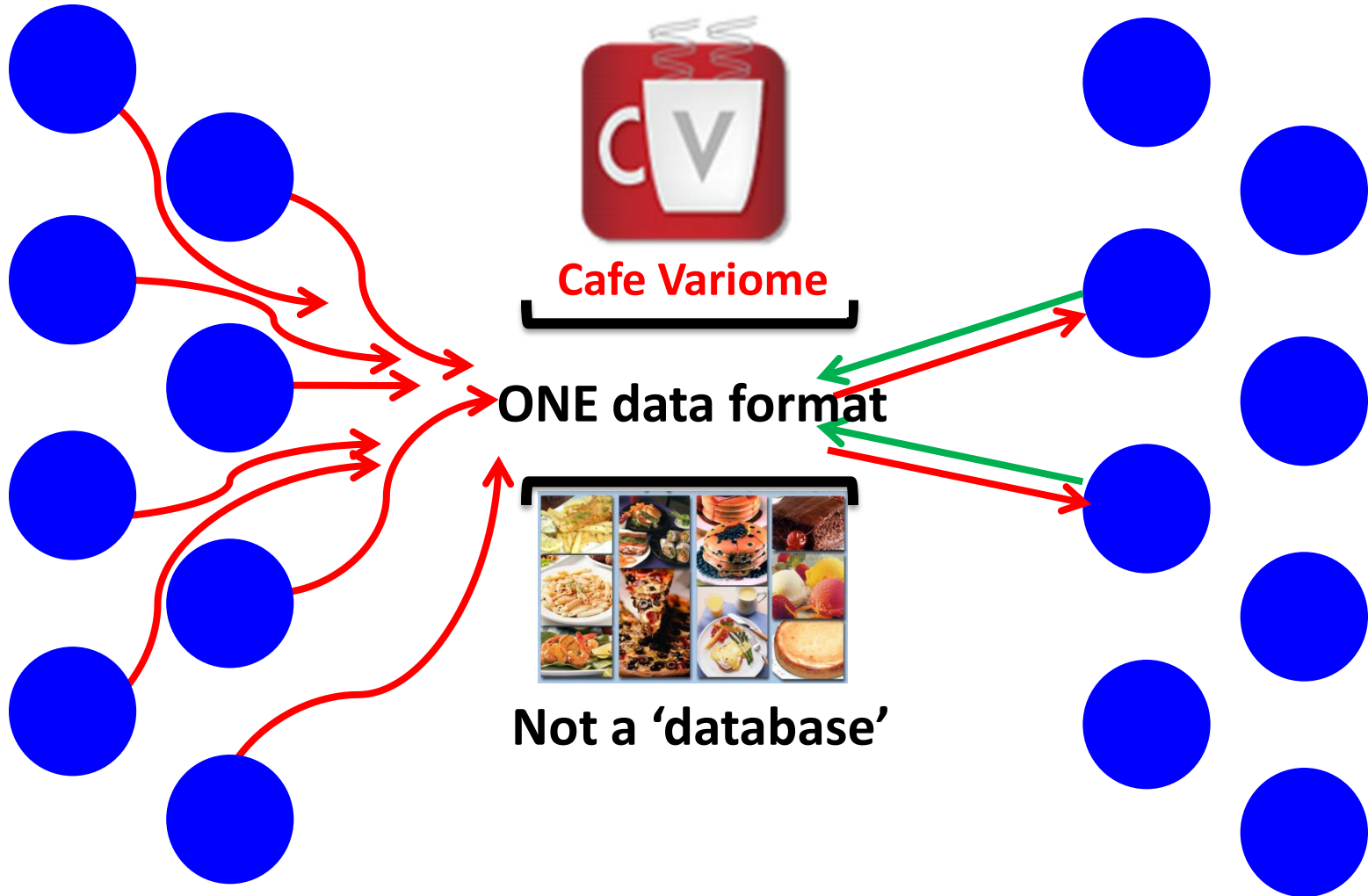
Mutation data from research groups such as
LSDBs, diagnostic research, and clinical miners/curators





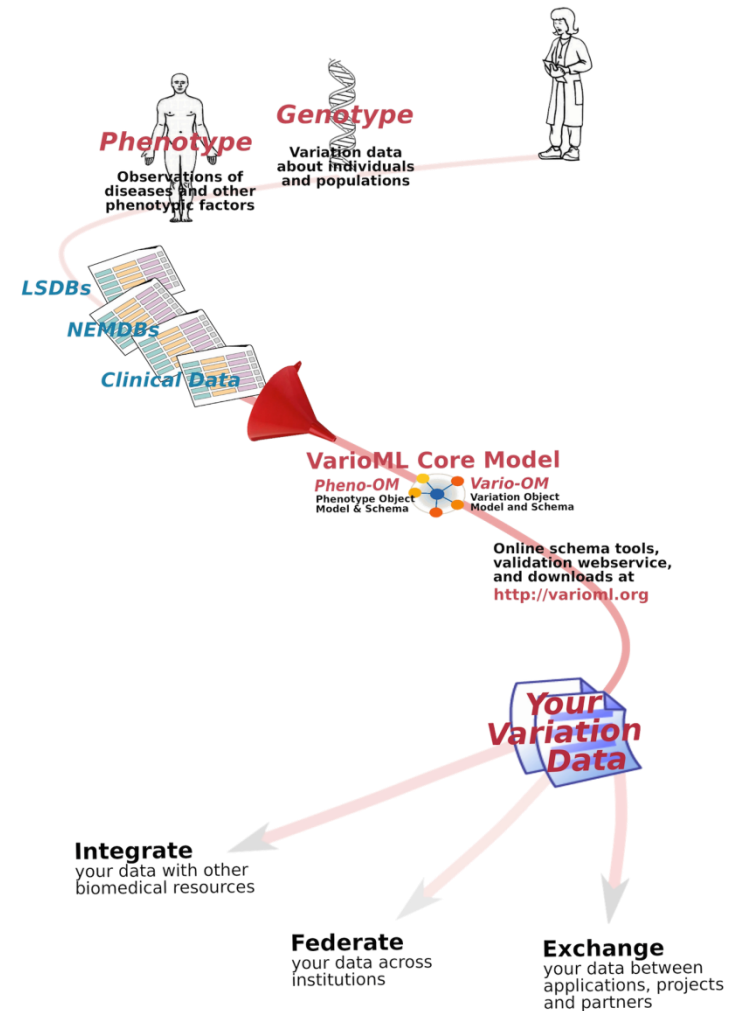
SUBMITTERS

USERS



VarioML

- XML format elements for LSDB data exchange use cases
 - Same format components for different applications
- Based on the Pheno-OM
 - Well defined semantics
- Intermediate format for semantic web
 - XSLT transformation to RDF
- Tools
 - Validators, JavaAPI, XSLTs



“Café Rouge enabled” Gensearch DNA analysis tool (Phenosystems)

The screenshot displays the Gensearch DNA analysis tool interface, specifically the Café Rouge validation panel. The panel is titled "Café Rouge validation panel" and contains several sections for user input and data display.

Information to be published via Café Rouge

- User identification:**
 - Laboratory ID: Lab ID
 - Submitter ID: Submitter ID
 - Author names: Doe J. Smith F. P.
- Data access control parameters:**
 - Sharing policy: Public
 - Embargo date: Month Year
 - Use restrictions: None
- Subject information:**
 - Anonymous subject ID: 1234
 - Subject type: Index case
 - Subject gender: Unspecified
 - Disease of interest: FH
 - HGNC gene name: LDLR
 - Comment: (empty text area)
- Variant list:**

HGVS name	Zygosity	Parental origin
c.1977C>A	On one sister autosome	Parent #1

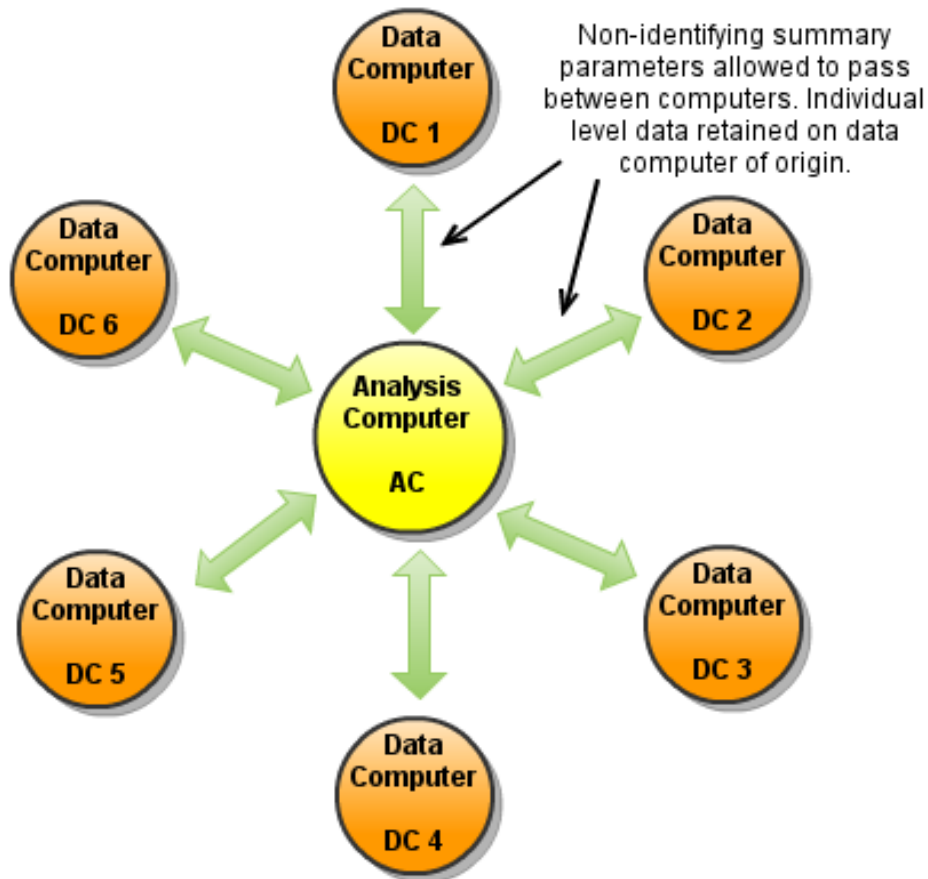
On the left side of the interface, there is a section for "Total number of potential mutation" and a "View" button. Below this, there are two gene models for EX13RLDL: 2178E13-RLDL_D12 and EX14RLDL: 2178E13-RLDL_D12, each with a 100% scale bar. At the bottom, there is a chromatogram showing the sequence A C A A C C T C A C C C A G.

A red arrow points from a text box to the "OK" button at the bottom of the panel.

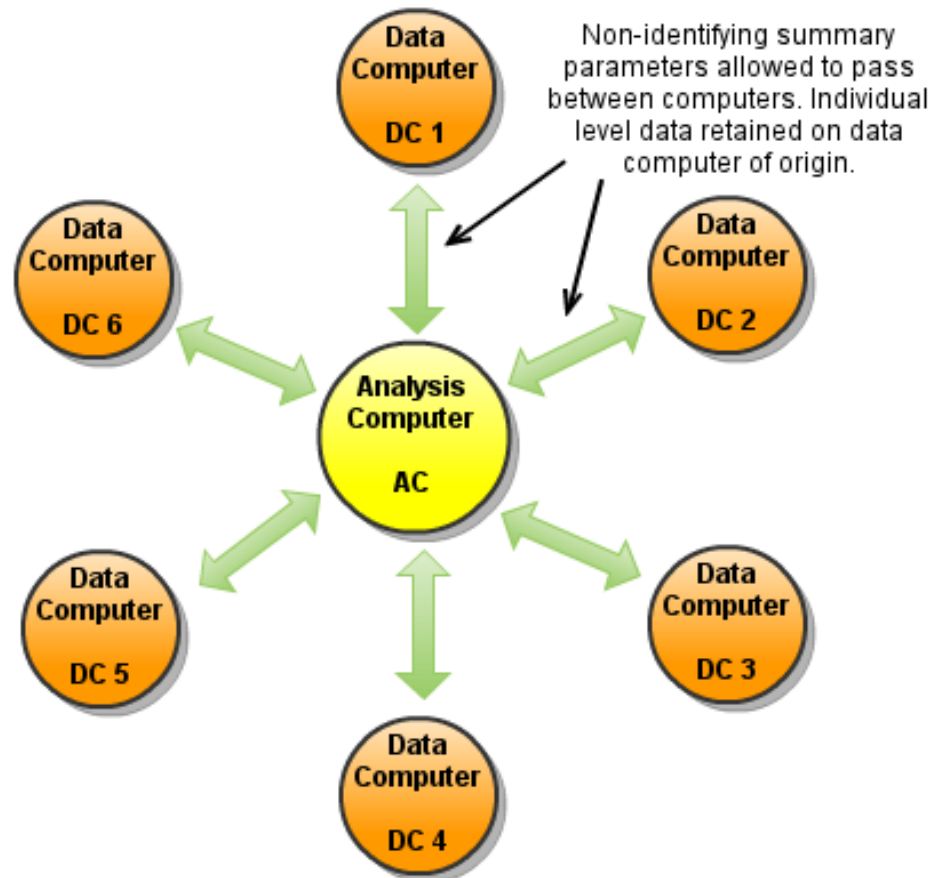
Uploaded via simple operation

DataSHIELD: Pooled data analysis without data sharing

- An Analysis Computer (AC) send iteratively requests for fitting a given GLM to the Data Computers (DC) on which data are stored



- Only summary statistics are sent back to the AC after each iteration
 - Individual-level data never leave DCs
- Eventually, iterations will converge to the **same** result as the model was fitted directly to the physically pooled data.



Local &/or Centralised &/or Federated technologies for data display and data mining

Solutions for open sharing:
summary level data, metadata,
& obfuscation strategies

Solutions for controlled sharing:
individual level data,
primary and/or harmonised data

Means for controlled and/or open
data use without sharing:
via DataShield

Eliminate ambiguity, maximise security, and enable recognition/reward:

- Digital IDs for scientific publications (DOIs)
- Digital IDs for Data Releases (DataCite)
- Digital IDs for Researchers (ORCID/OpenID)
- Digital IDs for BioResources (BRIF)

Tool for discovery of
sample collections + original + harmonised variables + counts/means

DataShaper
development and use

Web services

New database for
sample collections,
variables + results

Web services

Existing database for
sample collections,
variables + results

Web services

Existing database for
sample collections,
variables + results

Today's Healthcare

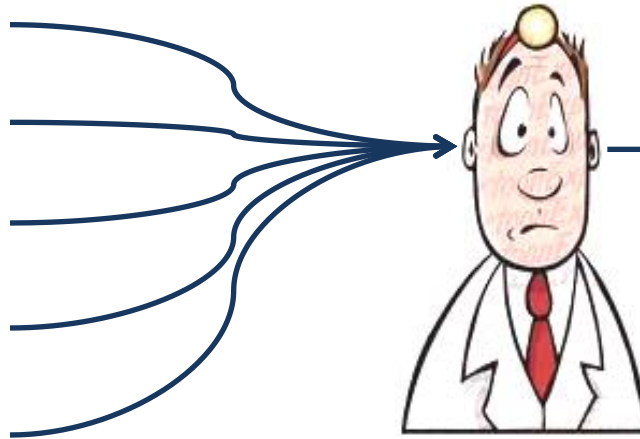
Medical Literature

Primary Research

Clinical Experience

Pharmacology

Diagnostics



**Inconsistent &
sub-optimal
health-care**

Tomorrow's Healthcare

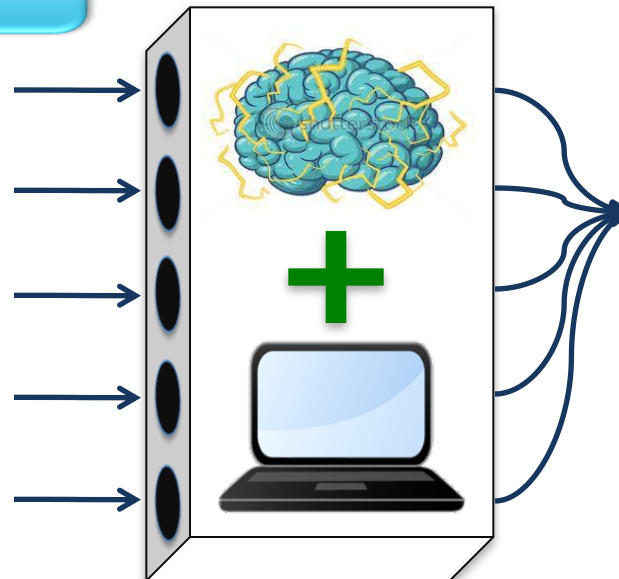
Medical Literature

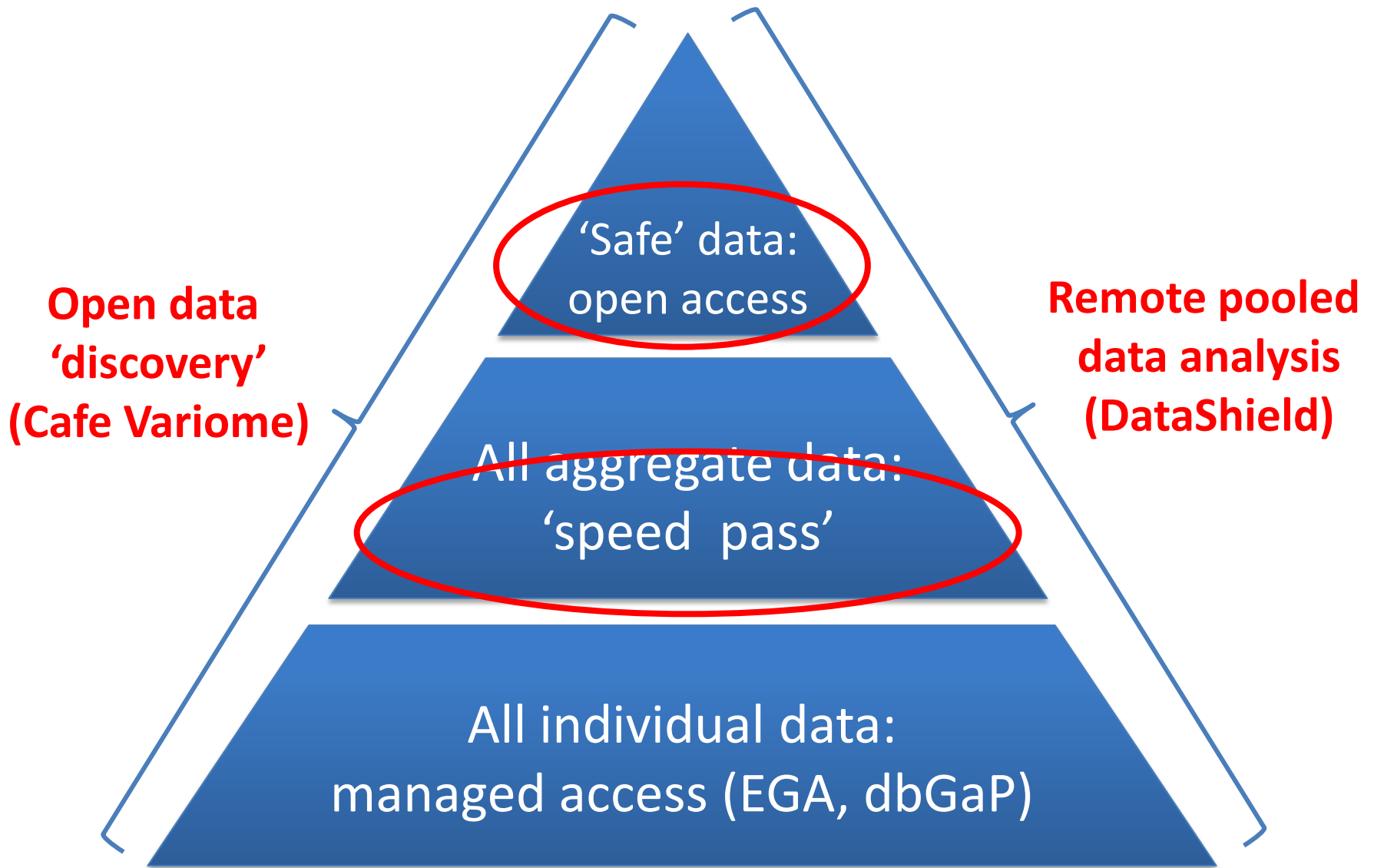
Primary Research

Clinical Experience

Pharmacology

Diagnostics





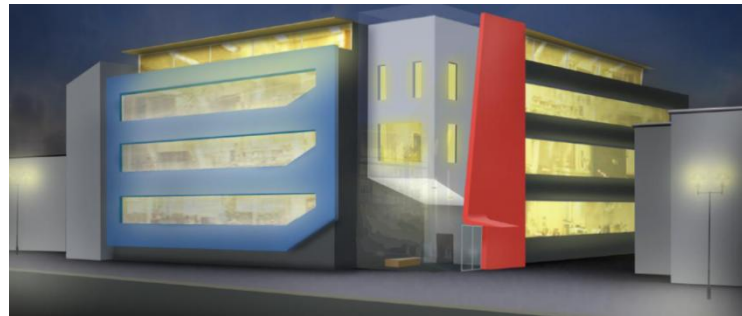
Issues that restrict sharing data

- Researchers may not have time nor funding to manually submit data, and/or submission process and requirements too complicated
- Researchers receive little or no recognition or reward for releasing data, hence little incentive to try
- Researchers may have positive reasons for NOT wanting to share data (ethical, legal, competitive edge)
- No current SANCTIONS for researchers that do not maximally share data

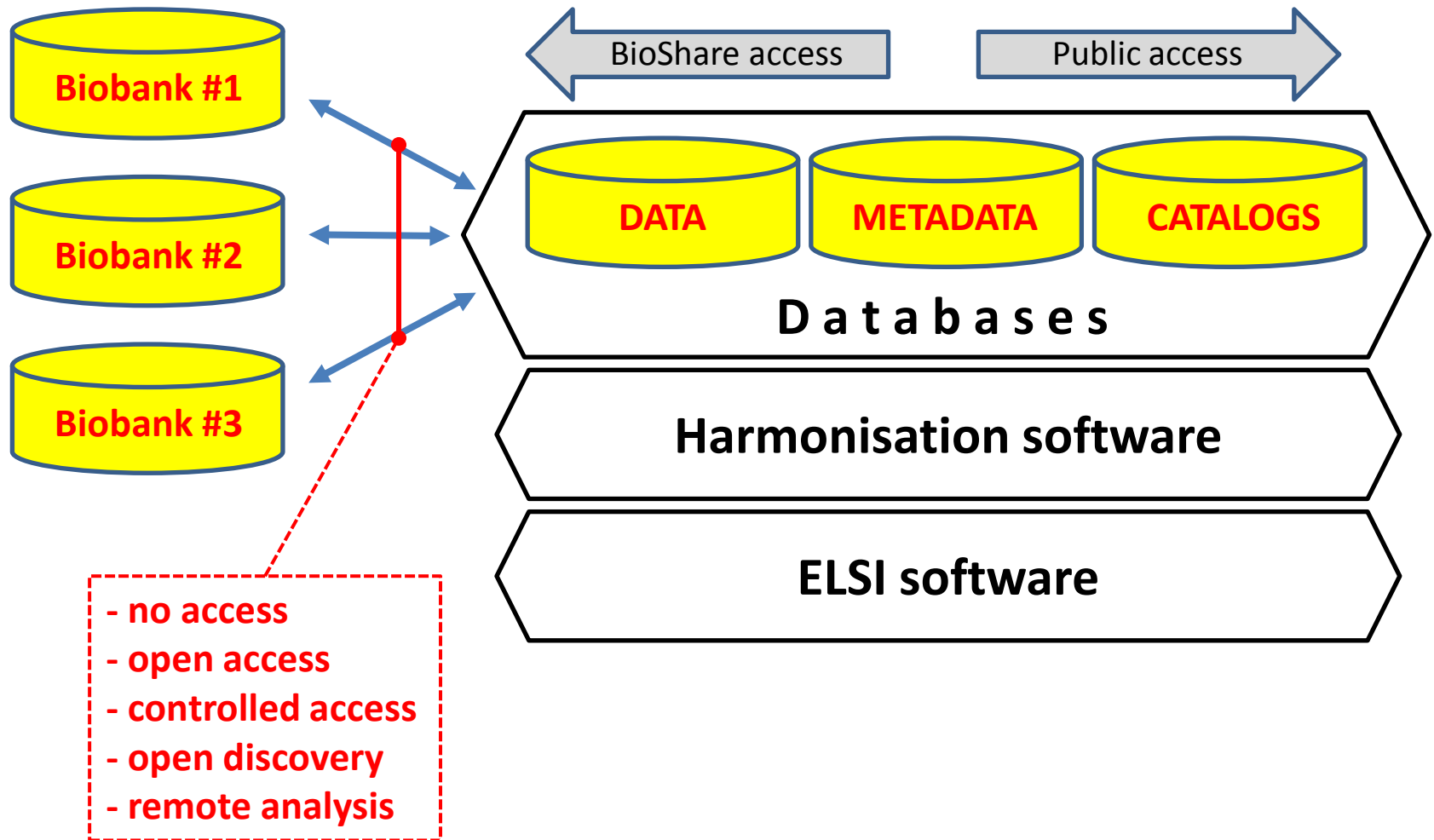
Acknowledgments

- GEN2PHEN Partners
- My team:
Robert Free, Rob Hastings, Adam Webb, Tim Beck, Sirisha Gollapudi, Gudmundur Thorisson, Owen Lancaster

**“Data-to-Knowledge-to-Practice”
(D2K2P) Center**

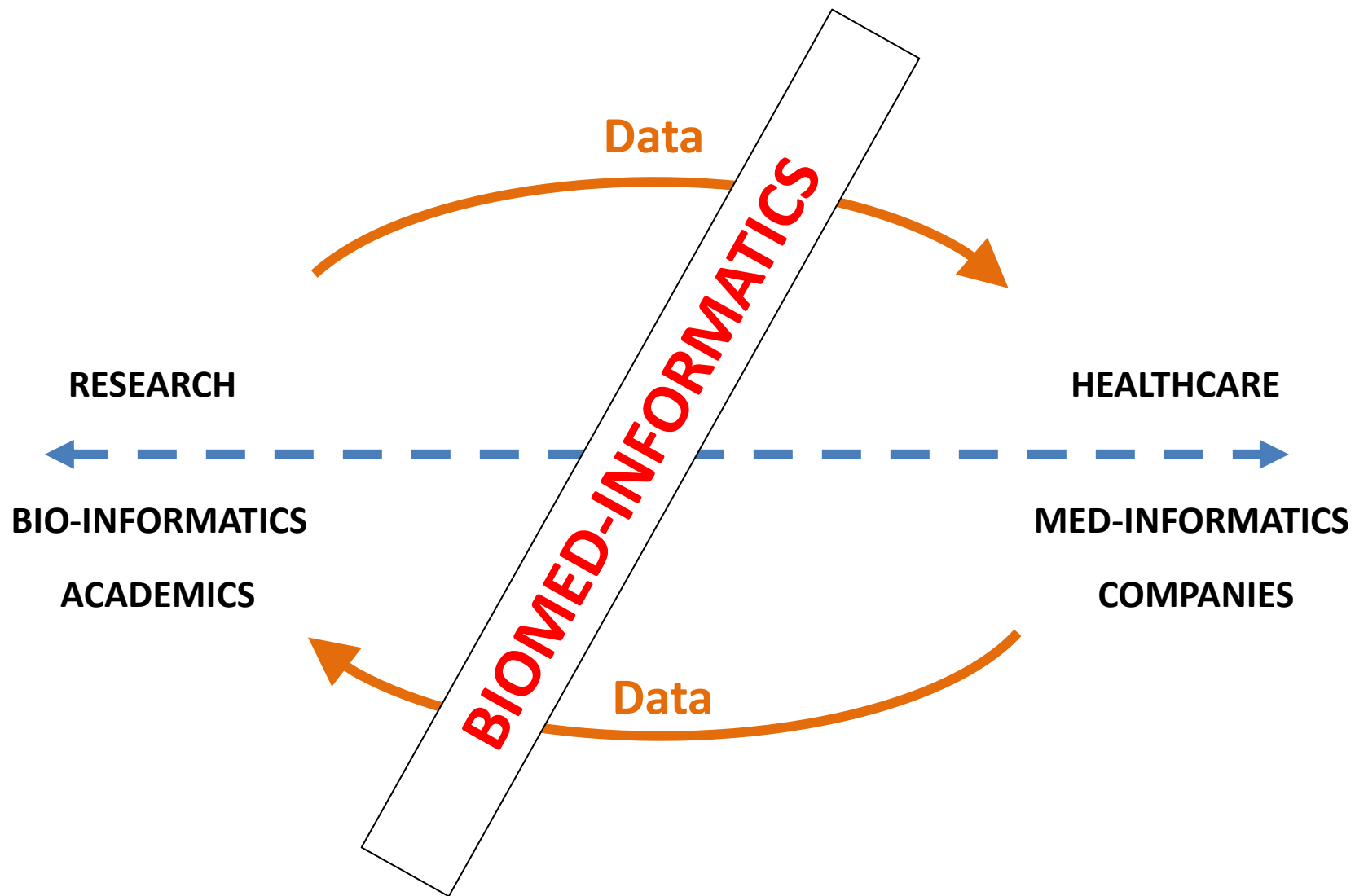


BIOBANKING ('BioShaRE')

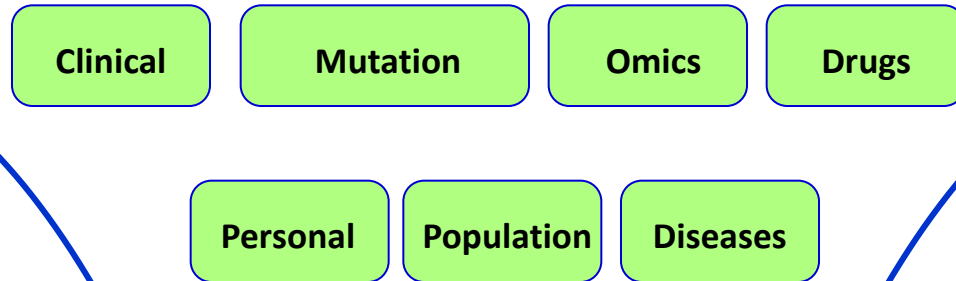


..... and / or

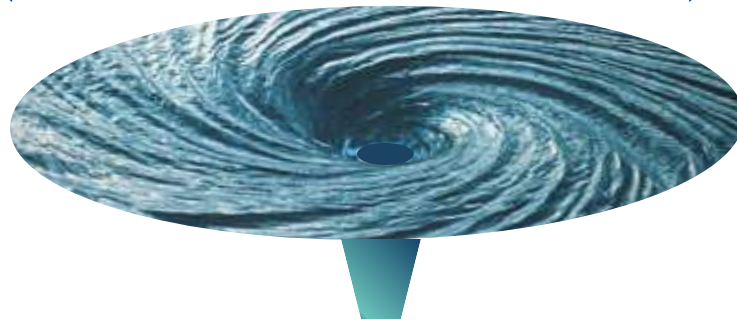
- **Open access (to any/all sensitive data) for data discovery purposes, without revealing data**
- **Open access (to any/all sensitive data) for pooled remote analysis**



**Data
+
Information
+
Knowledge**



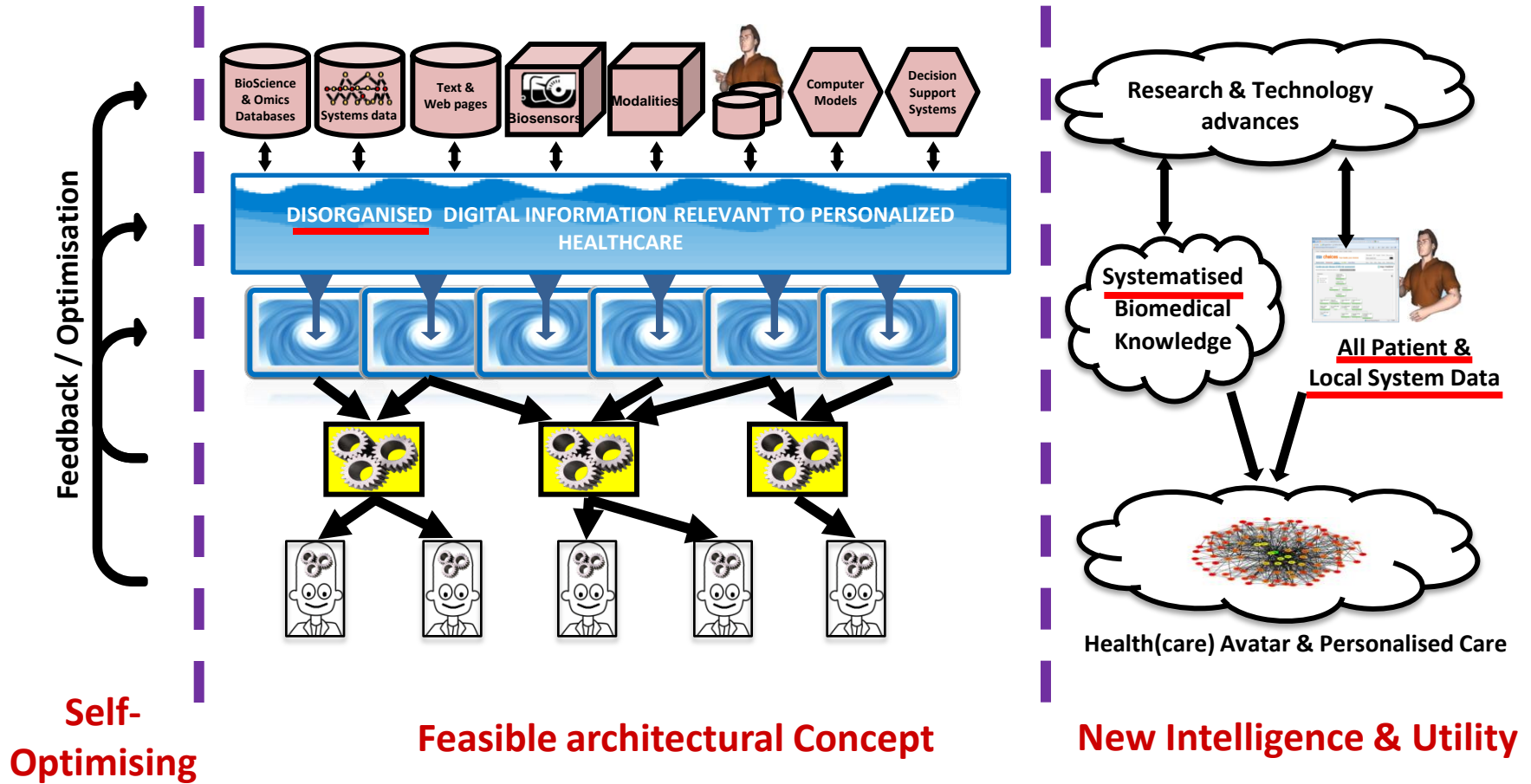
**Disease specific
Portals**



**Health Care
Utility**



The I-Health Opportunity



Progress to date:

- operating as part of GEN2PHEN extended goals
- created 'I-Health community', >150 academics, companies, healthcare providers
- concept presented in many international meetings and forums
- free 1/2 day workshop as satellite to ESHG (6 invited speakers, funding in place)
- major international conference in Brussels, Oct 2011 (venue booked, funding in place)
- organising a 3-day exploratory 'think tank' in spring 2012, with PHG
- high level lobbying with funders and policy makers
- incorporating I-Health elements in EUR 70M of funding applications due autumn 2011
- launching the Leicester D2K2P Center, to implement I-Health concepts

**“Data-to-Knowledge-to-Practice”
(D2K2P) Center**

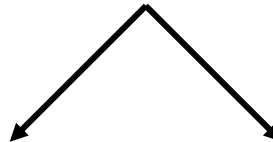


Issues related to GWAS data sharing

- **Researchers are not sharing G2P data generally for various reasons.....**
 - Insufficient staffing &/or bioinformatic capabilities
 - Ethical issues / identifiable data (genotypes, phenotypes) / privacy
 - Desire to monopolise and control “their data”
 - No credit/recognition is given for data sharing or curation
- **Lack of sharing is harming the scientific endeavor.....**
 - Most information not available to most researchers for consideration
 - Heterogeneity across studies/populations, and smaller effect sizes missed
 - Missed opportunities for collaboration & researcher recognition & reward

Identifying Individuals in Aggregated Data

AGGREGATE LEVEL DATA



Safe Elements:

- P values & odds ratios
 - graphically, all markers
 - non-directional, all markers
 - directional, hundreds of markers
- Allele freqs (hundreds of markers)

Open Access

Unsafe Elements:

- P values & odds ratios
 - directional, all markers
- Allele freqs
 - all markers

'Speedy' Access

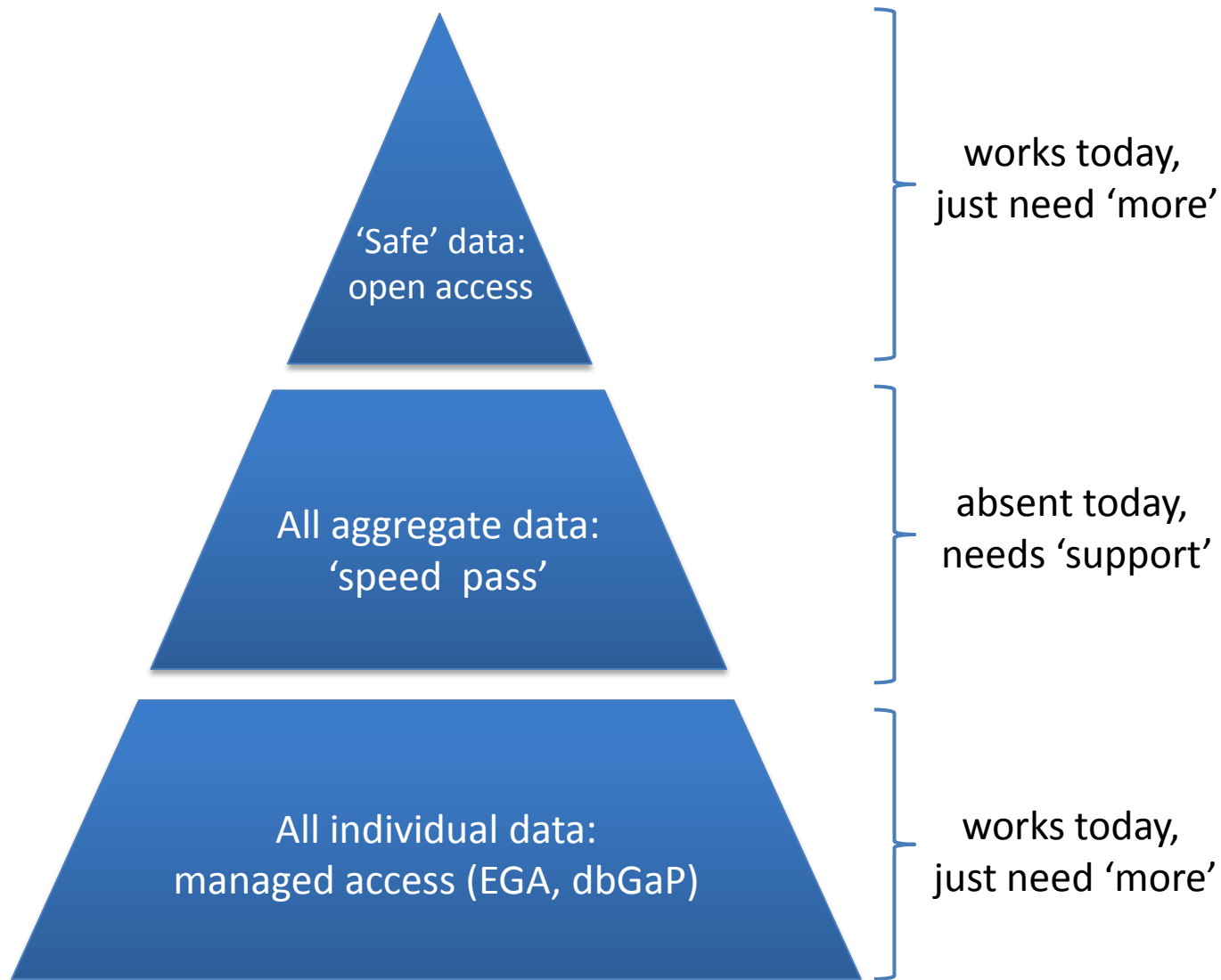
Solving issues in modern bioscience research relating to...

- researcher disambiguation
- data access control
- data sharing & online publication
- tracking & rewarding data contributions
- data integration & knowledge mining

...via people having Digital Identities on the web

DataSHIELD: Pooled data analysis without data sharing!!

- Conventionally, for individual-level analysis,
 - one pools the data from each of the studies into one single large dataset
 - Then, analyses this data set as it was a single study.
- Requires to have access to individual-level data
 - ELSI restriction on 3rd party sharing
- For a wide class of analyses (GLMs), this can be avoided using the DataSHIELD approach (Wolfson et al, *IJE* 2010)
- DataSHIELD can give same analysis results without disclosing any individual-level data to the researchers!



The journey to optimal data sharing...

Reluctance
to share

Ethico-legal
restrictions

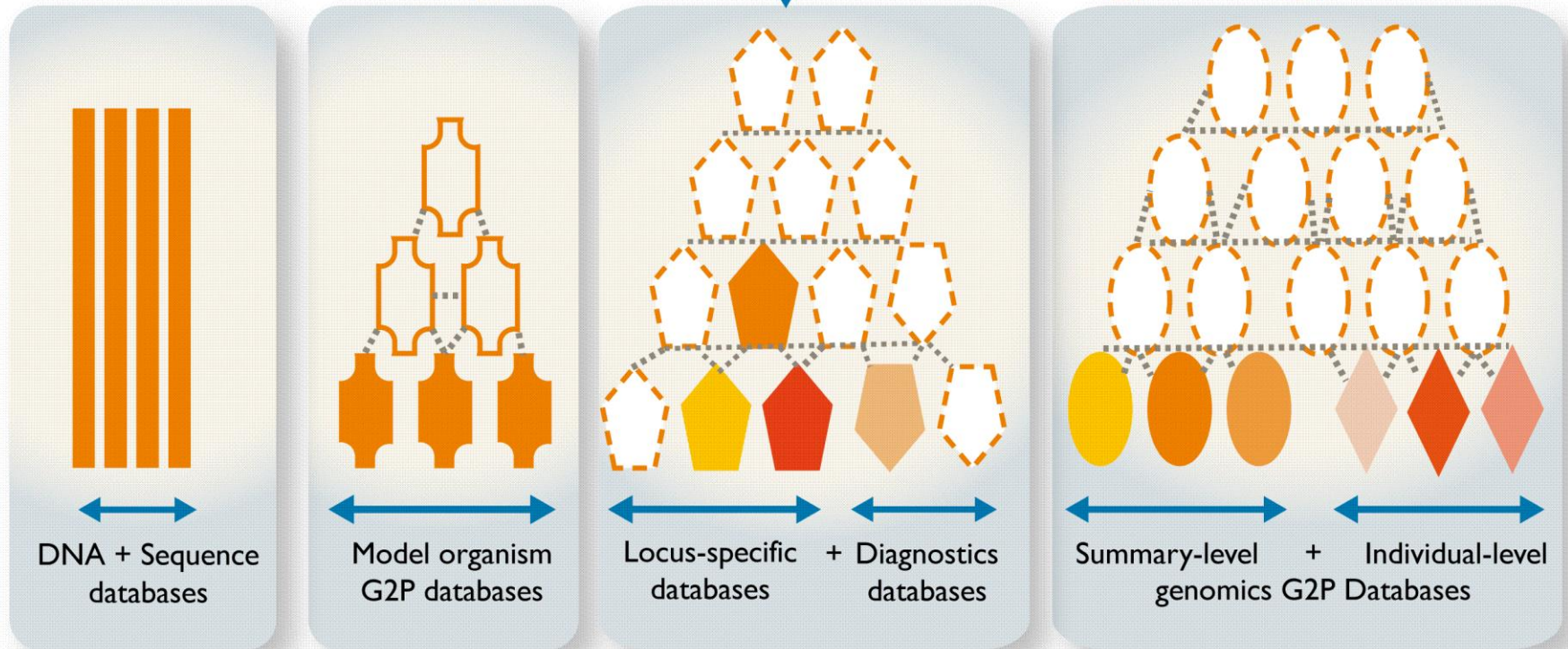
Technical obstacles
(integration, access, etc)

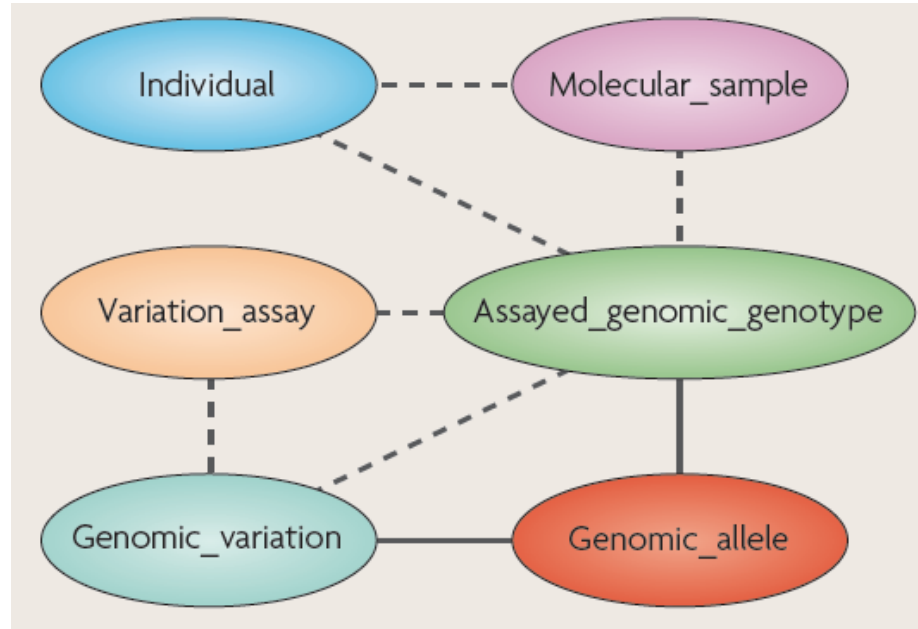


...tackle via people having Digital Identities on the web

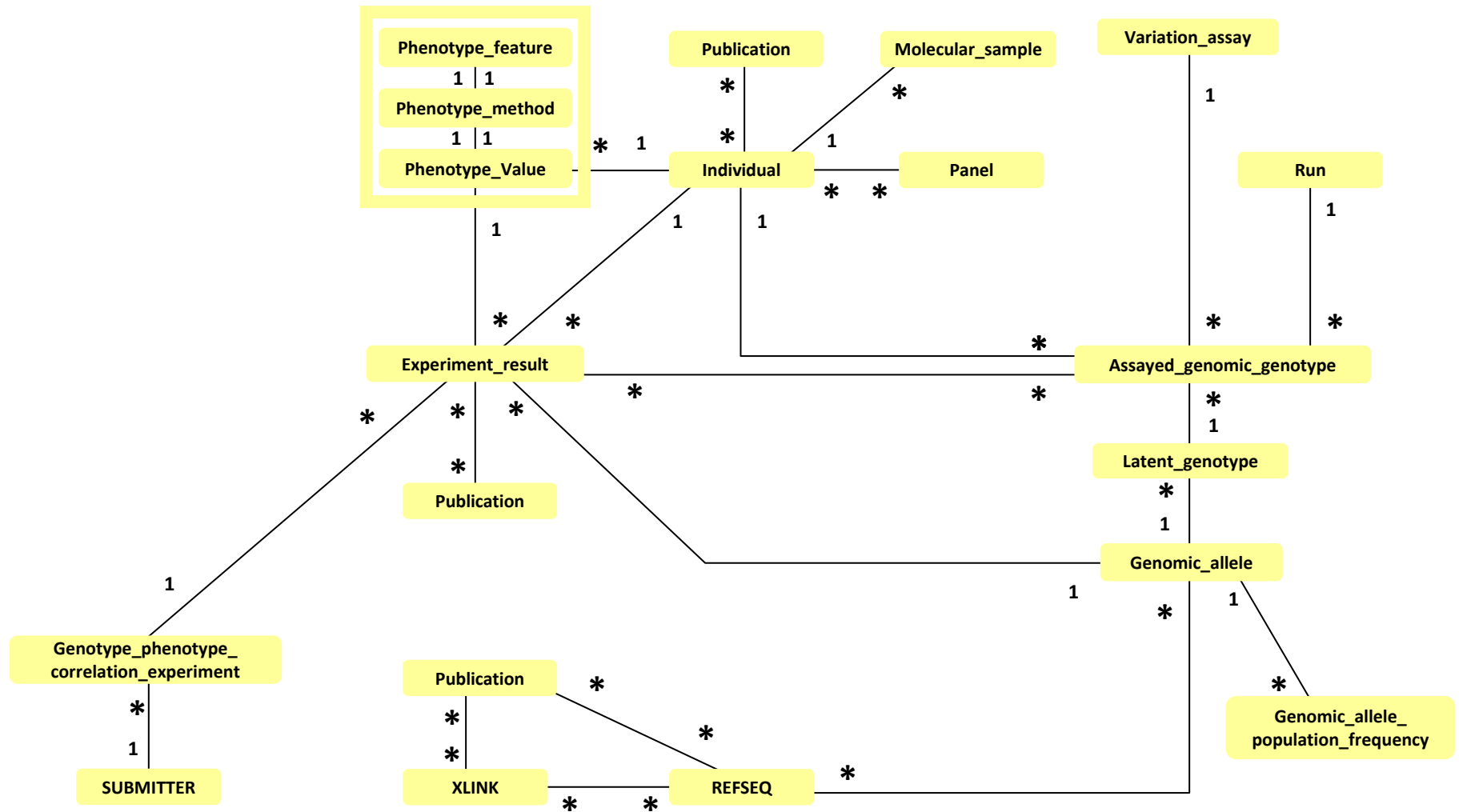
PUBLIC DOMAIN GENOME BROWSERS

e.g. Ensembl



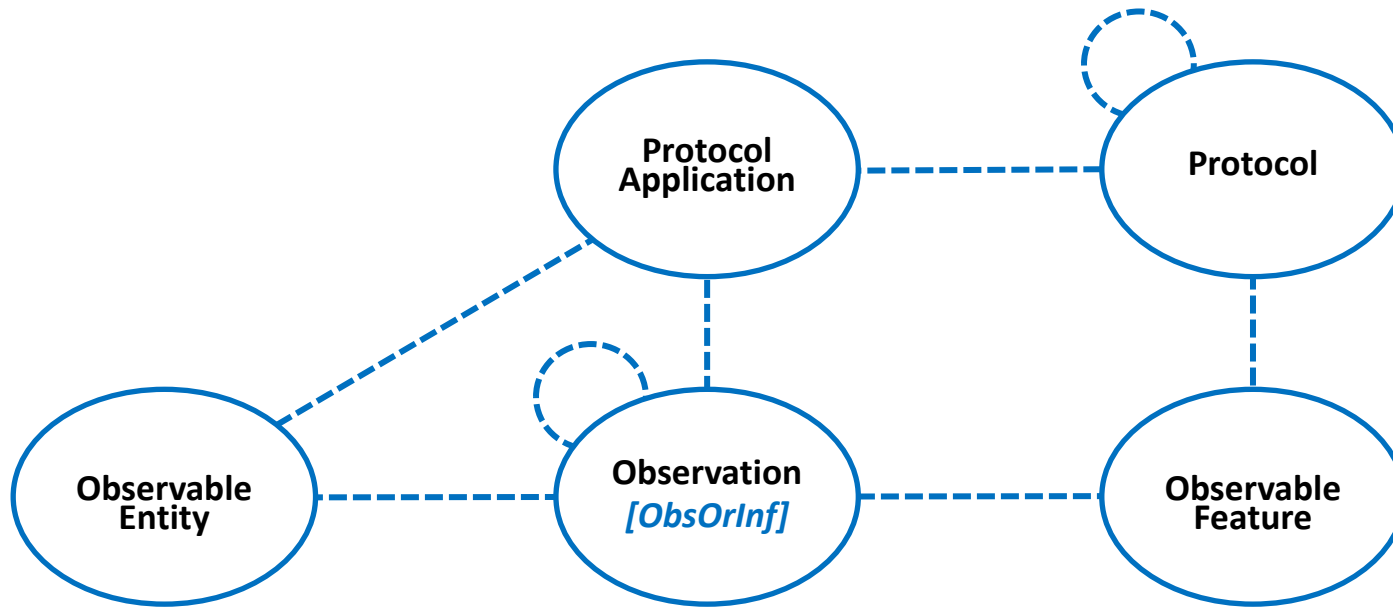


PaGE-OM
(Phenotype & Genotype Experiment
Object Model)



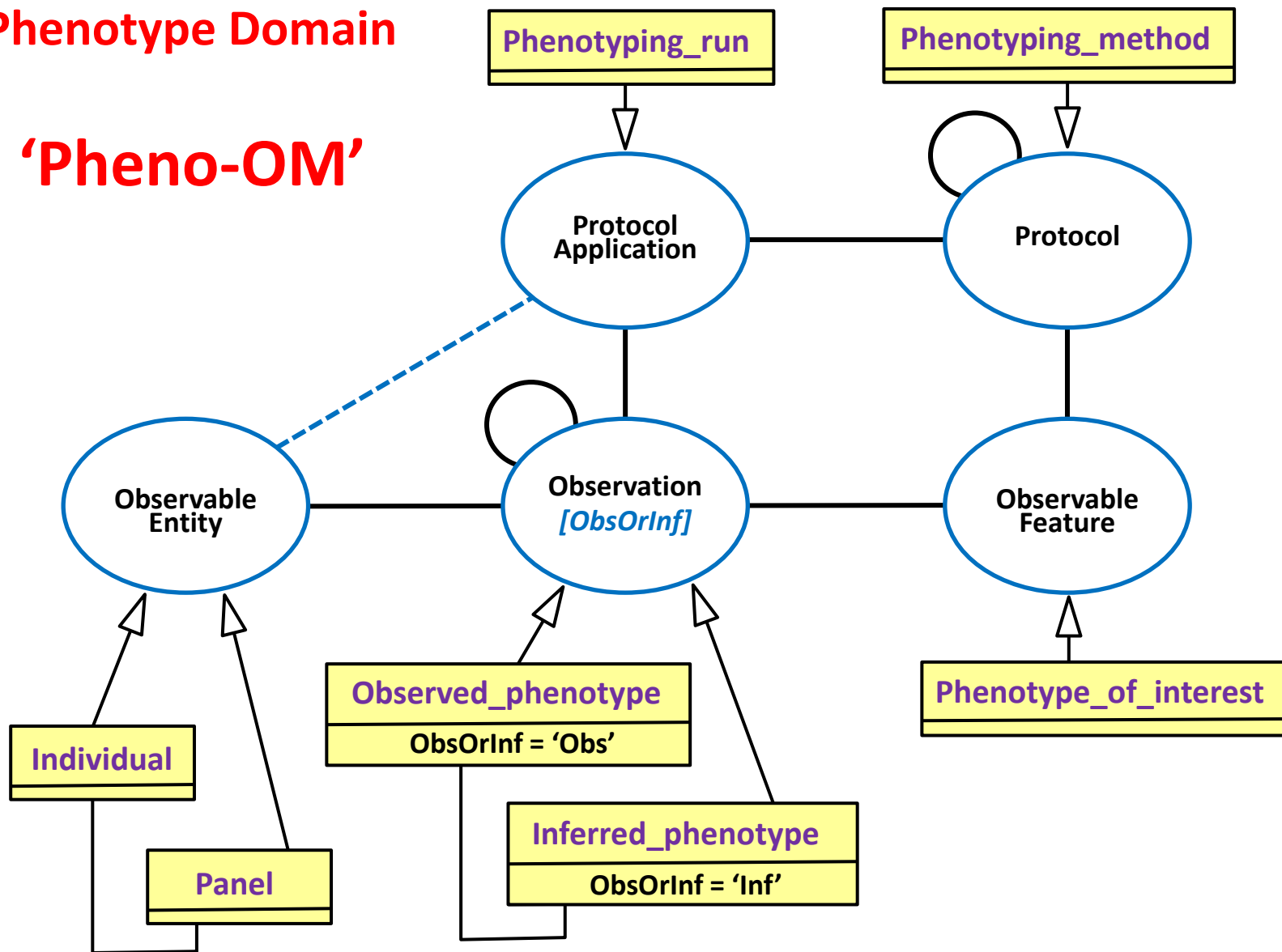
Universal, Core Data Model for LSDBs
(from LOVD, UMD, DMuDB, Findis)

Core Model



Phenotype Domain

'Pheno-OM'



Correspondence

Open access

Locus Reference Genomic sequences: an improved basis for describing human DNA variants.

Raymond Dalglish, Paul Eilander, Fiona Cunningham, Alex Astashyn, Raymond E Tully, Glenn Proctor, Yuan Chen, William M McLaren, Pontus Larsson, Brendan W Vaughan, Christophe Beroud, Glen Dobson, Heikki Lehtasalo, Peter EM Taschner, Johan T den Dunnen, Andrew Devereau, Ewan Birney, Anthony J Brookes and Donna R Maglott

For all author emails, please [log on](#).

Genome Medicine 2010, **2**:24 doi:10.1186/gm145
Published: 15 April 2010

Abstract (provisional)

As our knowledge of the complexity of gene architecture grows, and we increase expression, the process of accurately describing disease-causing gene variants is due to current reference DNA sequence formats that do not fully meet the Reference Genomic (LRG) sequence format which has been designed for this format builds on the successful National Center for Biotechnology Information single-file record containing a uniquely stable reference DNA sequence alone essential to the description of gene variants. In principle, LRGs can be created we recognise the need to respect legacy numbering systems for exons and these. We hope that widespread adoption of LRGs -- which will be created at the Bioinformatics Institute (EBI) -- along with consistent use of the Human Genome nomenclature will reduce errors in the reporting of variants in the literature affecting human health. Further information can be found on the LRG web site

EDITORIAL

nature
genetics

Conventional wisdom

Recent agreement on stable reference sequences for reporting human genetic variants now allows us to mandate the use of the allele naming conventions developed by the Human Genome Variation Society.

By agreement between stakeholders and two principal databases, it has been proposed (R. Dalglish *et al.*, *Genome Med.* 2, 24, 2010, doi:10.1186/gm145) that human genetic variants be reported relative to a new set of stable reference sequences, "Locus Reference, Genomic" (LRG, pronounced "large" <http://www.lrg-sequence.org/page.php>). These sequences have been developed from the initial NCBI RefSeqGene concept and are provided by NCBI and EBI according to agreed rules and in consultation with community users of locus-specific genetic information and locus-specific databases. It is anticipated that the LRG will be stable and supported for many years, long enough to serve as a bridge between existing and future clinical gene tests

age, resequencing and marker association studies and so keep allele descriptions commensurate with the method by which their data were generated.

The LRG reference sequences should be used in conjunction with standard HGNC gene abbreviations (<http://www.genenames.org/>) that we already require as a condition of publication. All human genetic variants must now be described—in abstracts and at first use—in accordance with the Human Genome Variation Society (HGVS) conventions (<http://www.hgvs.org/mutnomen/>) also as a condition of publication. We continue to encourage authors to use HGVS nomenclature for unambiguous reference in all tables and figures and throughout the

PUBLIC DOMAIN GENOME BROWSERS

e.g. Ensembl

- ! over 2000 standardised & interoperable LSDBs
- ! Web-services on top of these databases
- ! merging & centralisation of summary contents
- ! comprehensive listing of all LSDBs (with HGVS/HVP)

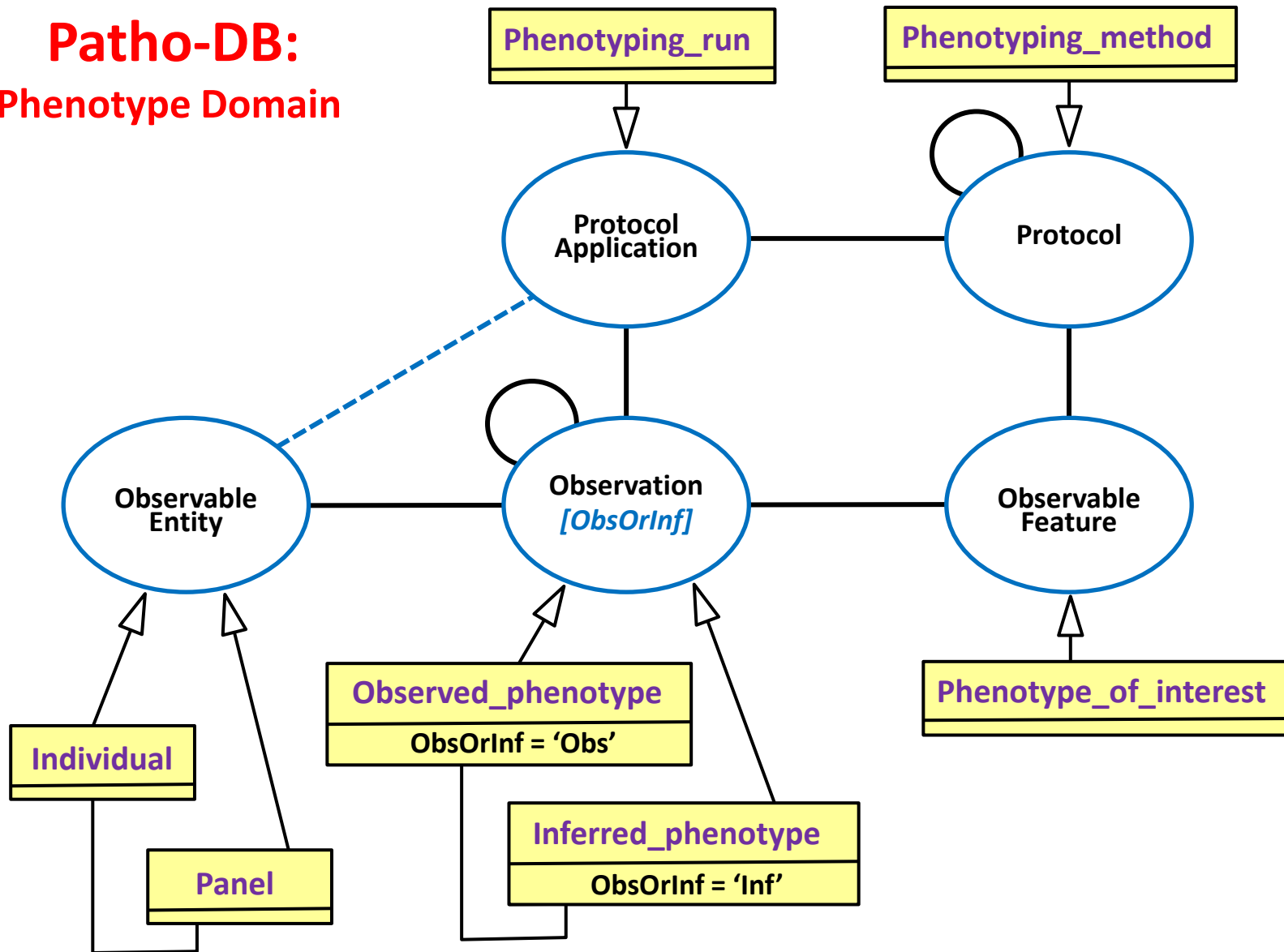
DNA + Sequence
databases

Model organism
G2P databases

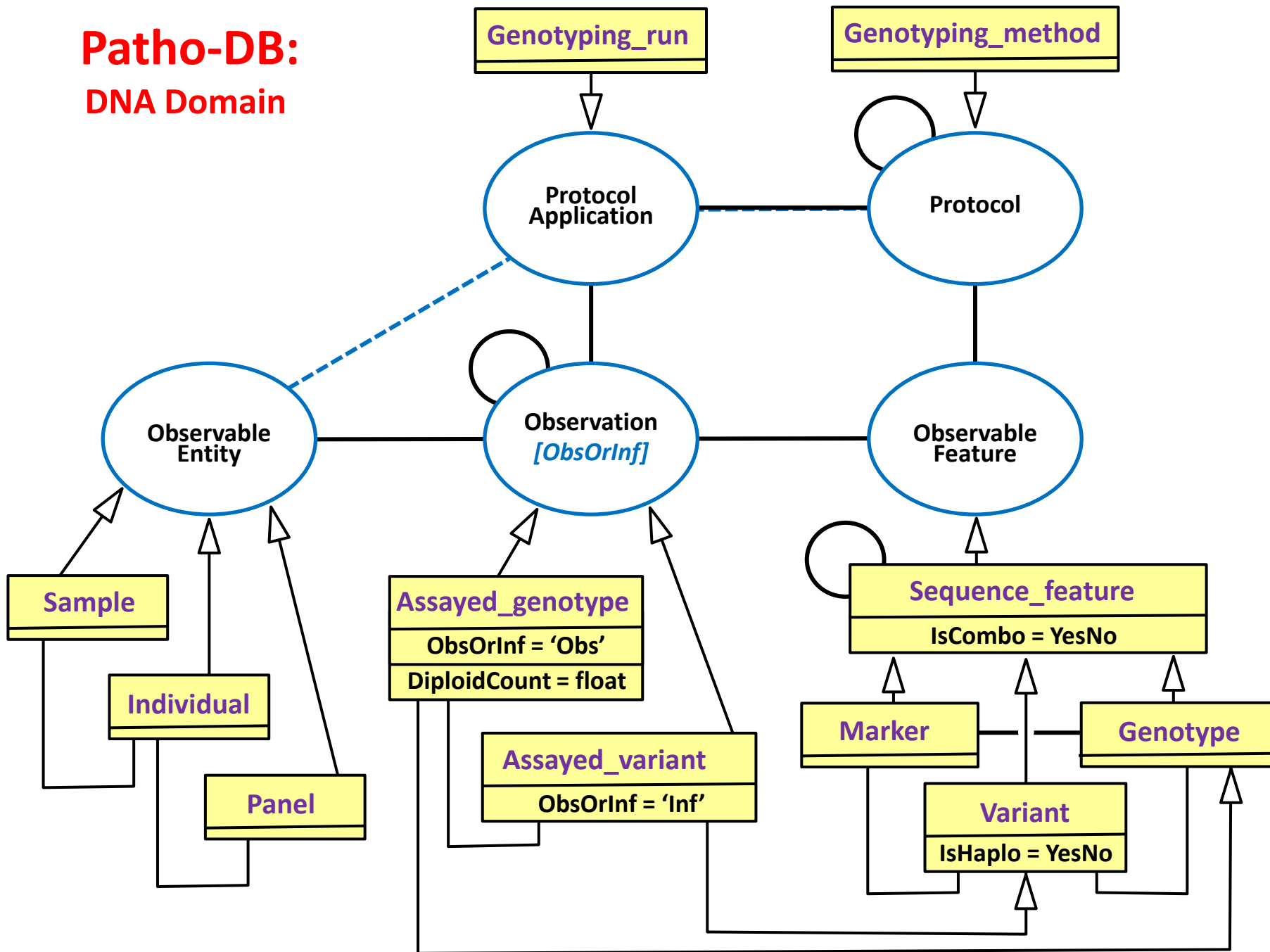
Locus-specific
databases + Diagnostics
databases

Summary-level
genomics G2P Databases + Individual-level
G2P Databases

Patho-DB: Phenotype Domain



Patho-DB: DNA Domain



PROJECTS:

GEN2PHEN

technologies, standards, software, databases & policies towards seamless/holistic organisation and utility of Genotype-To-Phenotype information

BioShaRE-EU

Harmonization, standardization, implementation & utilization of biobanking research tools (sampling, computing & analysis technologies)

COPD-MAP

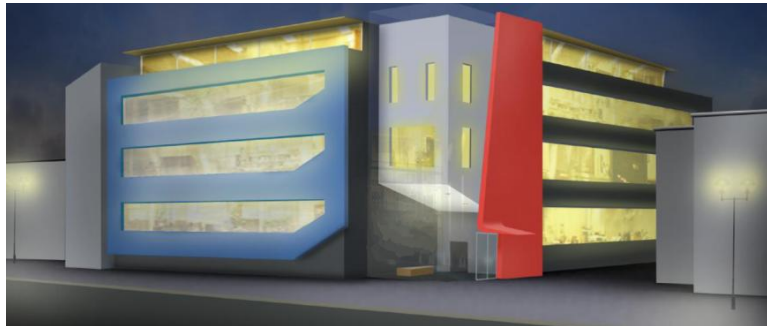
In charge of data management for £7M UK systems biology study into COPD. Exploring several platform options, including TransMart

'I-Health' Concepts

Mapping medical informatics needs to bridge the gap between research & healthcare informatics, part of the IT Future of Medicine Pilot being run by Hans Lehrach

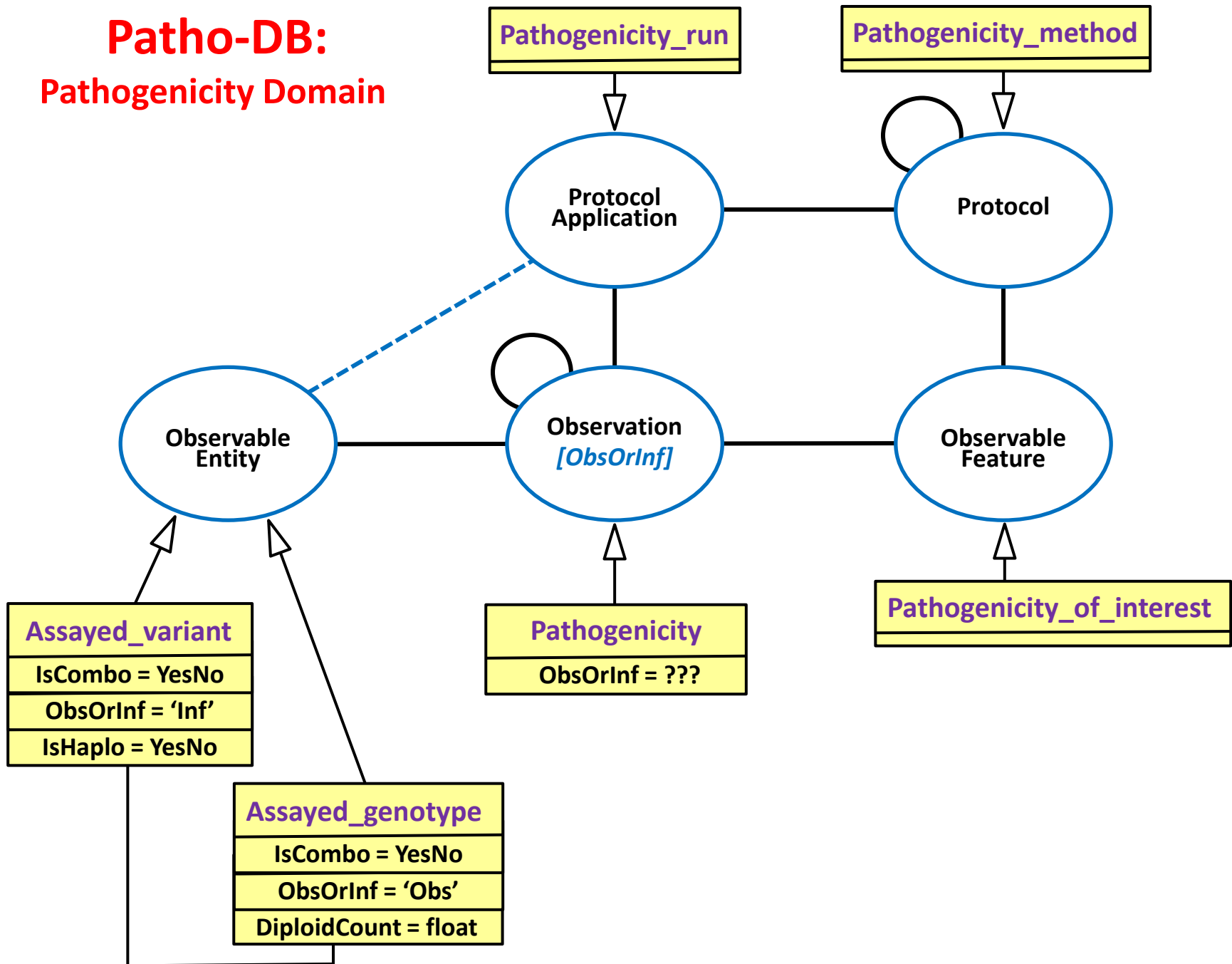
Data-2-Knowledge-2-Practice Centre (Director)

Two floors of biobank & I-Health IT, atop a CVD & respiratory disease clinic PLUS advanced biobank



Patho-DB:

Pathogenicity Domain



Enter a study or marker identifier, keywords, a gene name, or chromosomal region.

(e.g. [BRCA1](#), [chr12:13234...4534534](#), [12p13.33](#), [cancer](#), [rs2317951](#))



Studies

- *genetic association database*
- *integrates many ('all') datasets*
- *summary level data only*
- *links to data sources for primary data*



Download



About H

The Human Genome information (HGVbaseG2P) summary level findings and small. We actively projects, and encourage community. [See more..](#)

[Mailing list](#) or [RSS](#)

date [Read](#)

base 8 [Read](#)

2010/08/05

HGVbaseG2P releases study database 7 [Read](#)



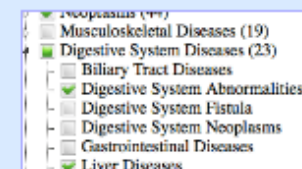
Frequently Asked Questions

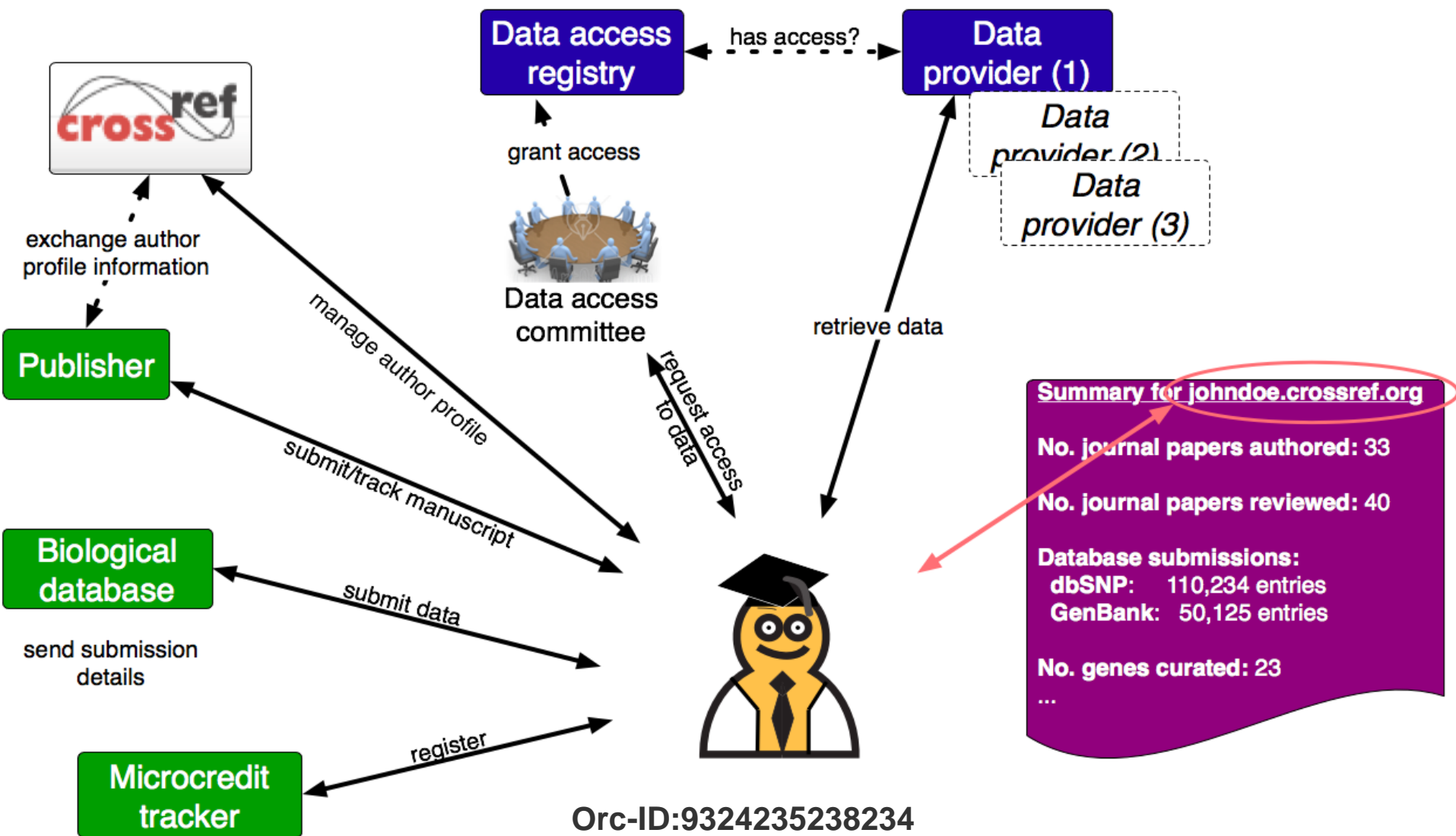
- How is the database content organised?
- How do I find Studies of interest?
- How do I find Markers of interest?
- How do I use the Browser to identify regions of interest?
- How do I submit my own data to HGVbaseG2P



Phenotype Trees

MeSH and HPO based [Phenotype trees](#) allow you to easily find Studies that relate to a specific disease or type of disease.





Orc-ID:9324235238234

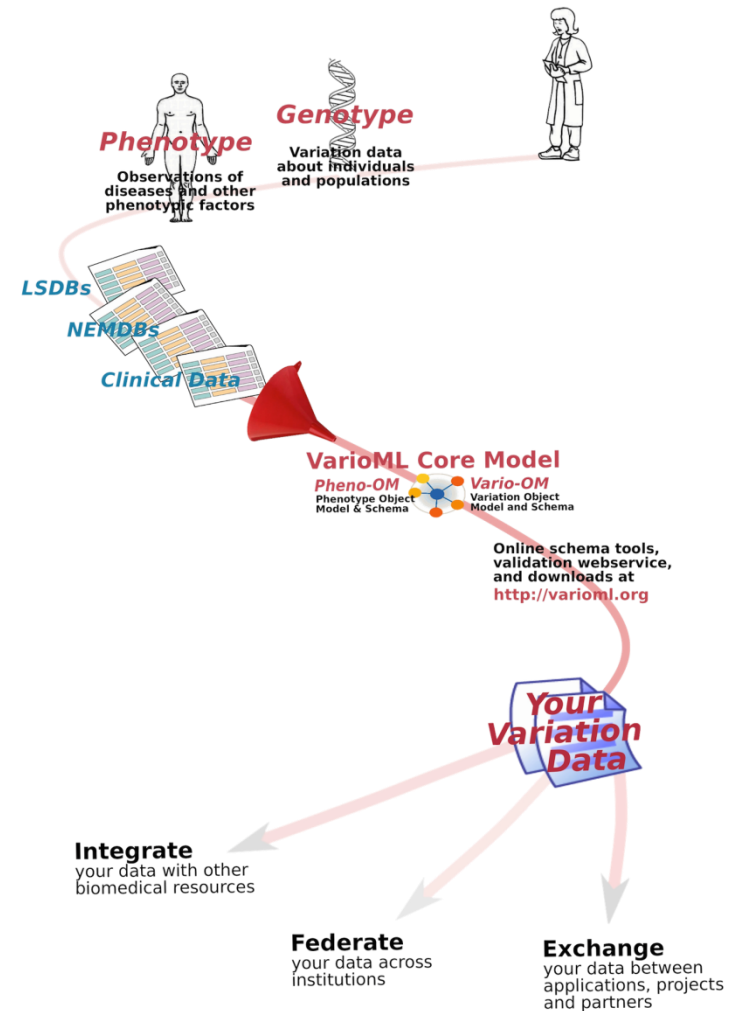
G. Thorisson, Univ. Leicester

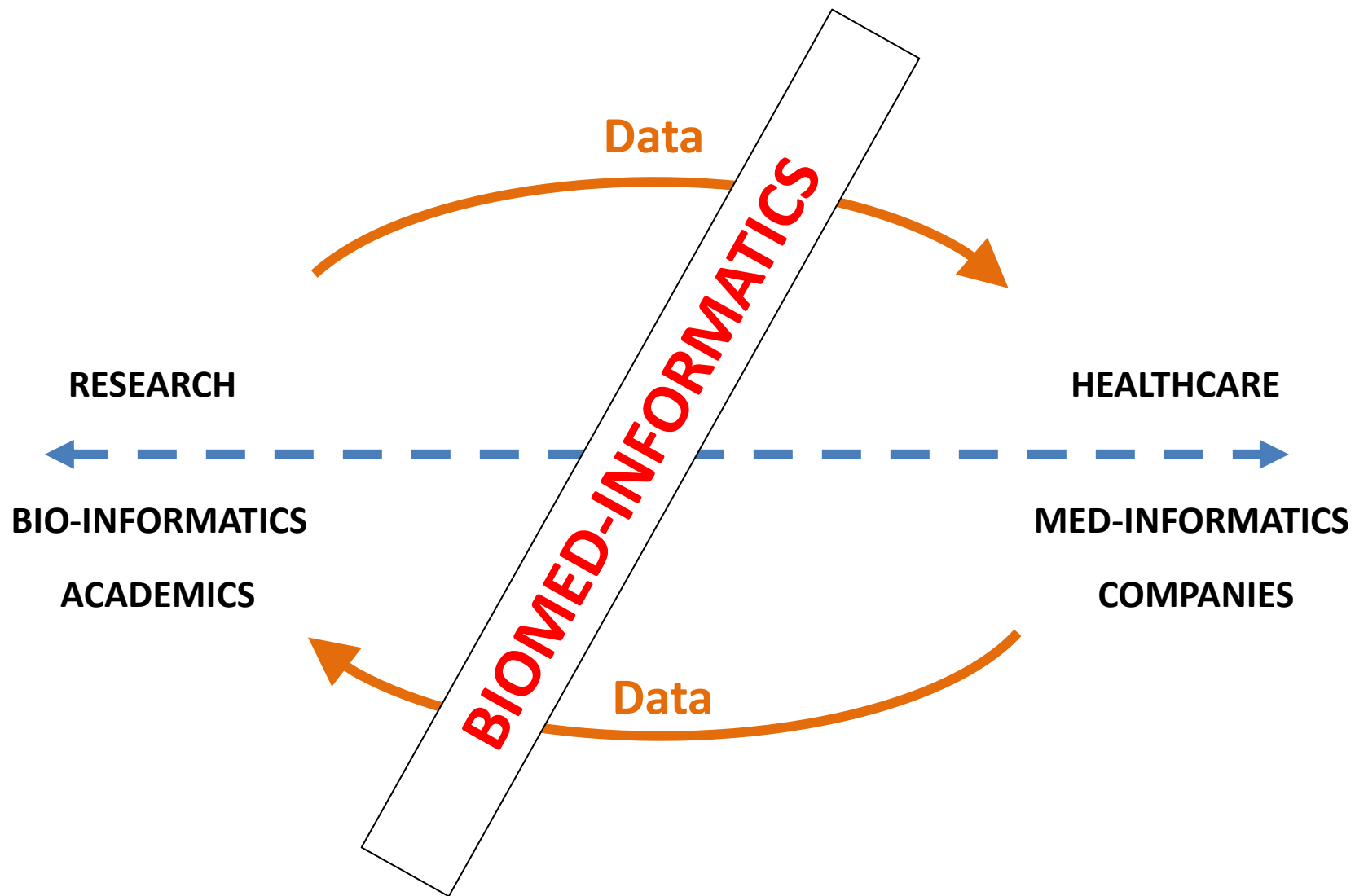
G. A. Thorisson, Univ. Leicester

G. A. Thorisson, Cold Spring Harbor Lab.

VarioML

- XML format elements for LSDB data exchange use cases
 - Same format components for different applications
- Based on the Pheno-OM
 - Well defined semantics
- Intermediate format for semantic web
 - XSLT transformation to RDF
- Tools
 - Validators, JavaAPI, XSLTs





Today's Healthcare

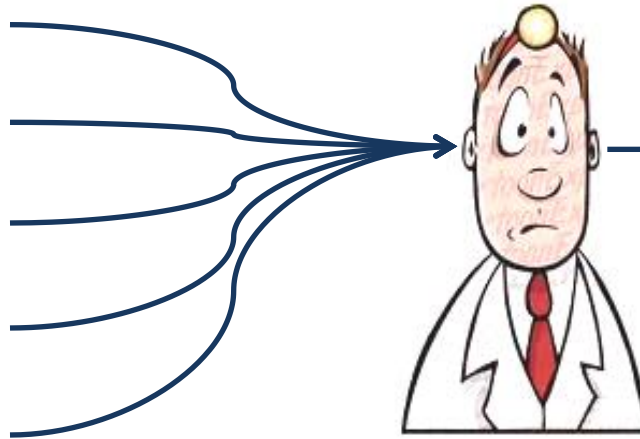
Medical Literature

Primary Research

Clinical Experience

Pharmacology

Diagnostics



**Inconsistent &
sub-optimal
health-care**

Tomorrow's Healthcare

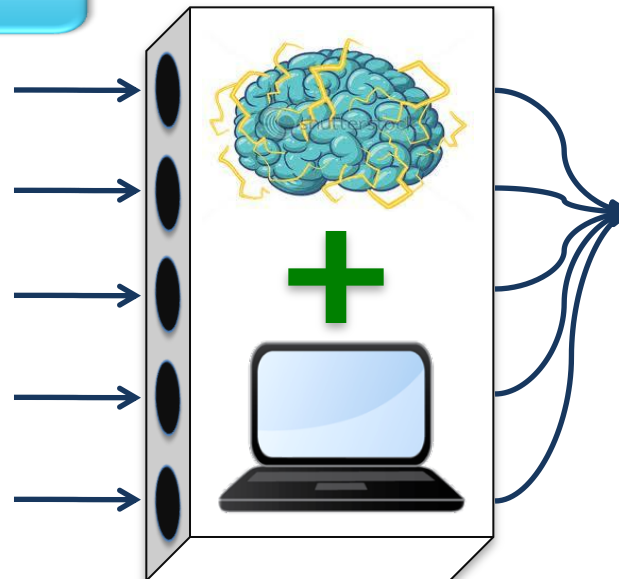
Medical Literature

Primary Research

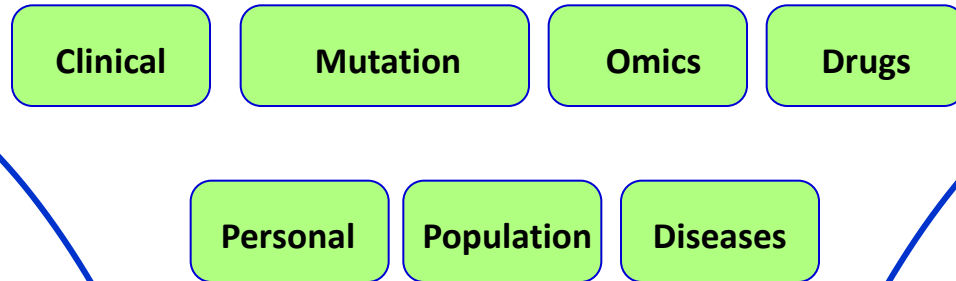
Clinical Experience

Pharmacology

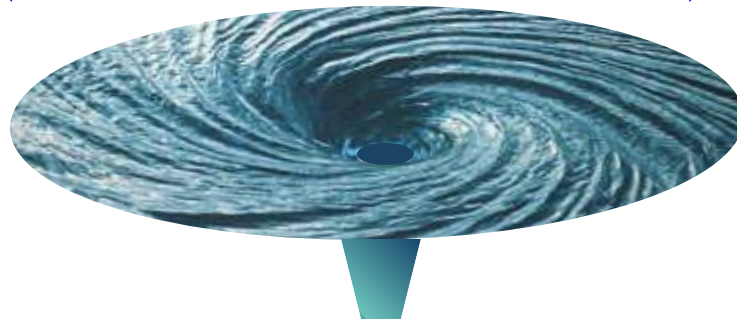
Diagnostics



**Data
+
Information
+
Knowledge**



**Disease specific
Portals**



**Health Care
Utility**

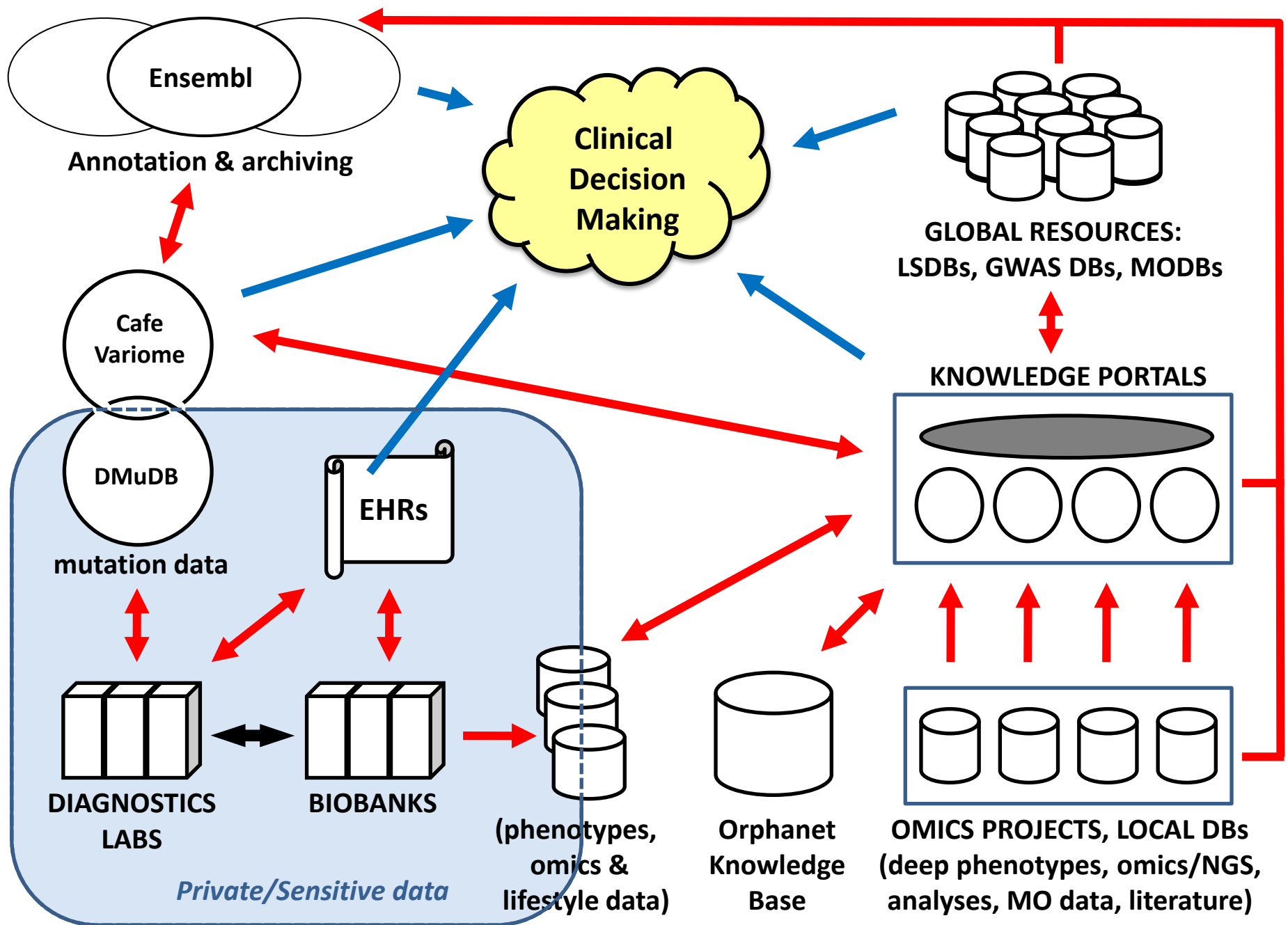


Acknowledgments

- GEN2PHEN Partners
- My team: *Robert Free, Rob Hastings, Adam Webb, Tim Beck, Sirisha Gollapudi, Gudmundur Thorisson, Owen Lancaster*
- I-Health supporters: *Iain Buchan, Barend Mons, Allan Hanbury, Jane Kaye, Hans Lehrach, Kurt Zatloukal, Jaak Vilo, Alvis Brazma, Carlos Diaz, + 150 other groups.*

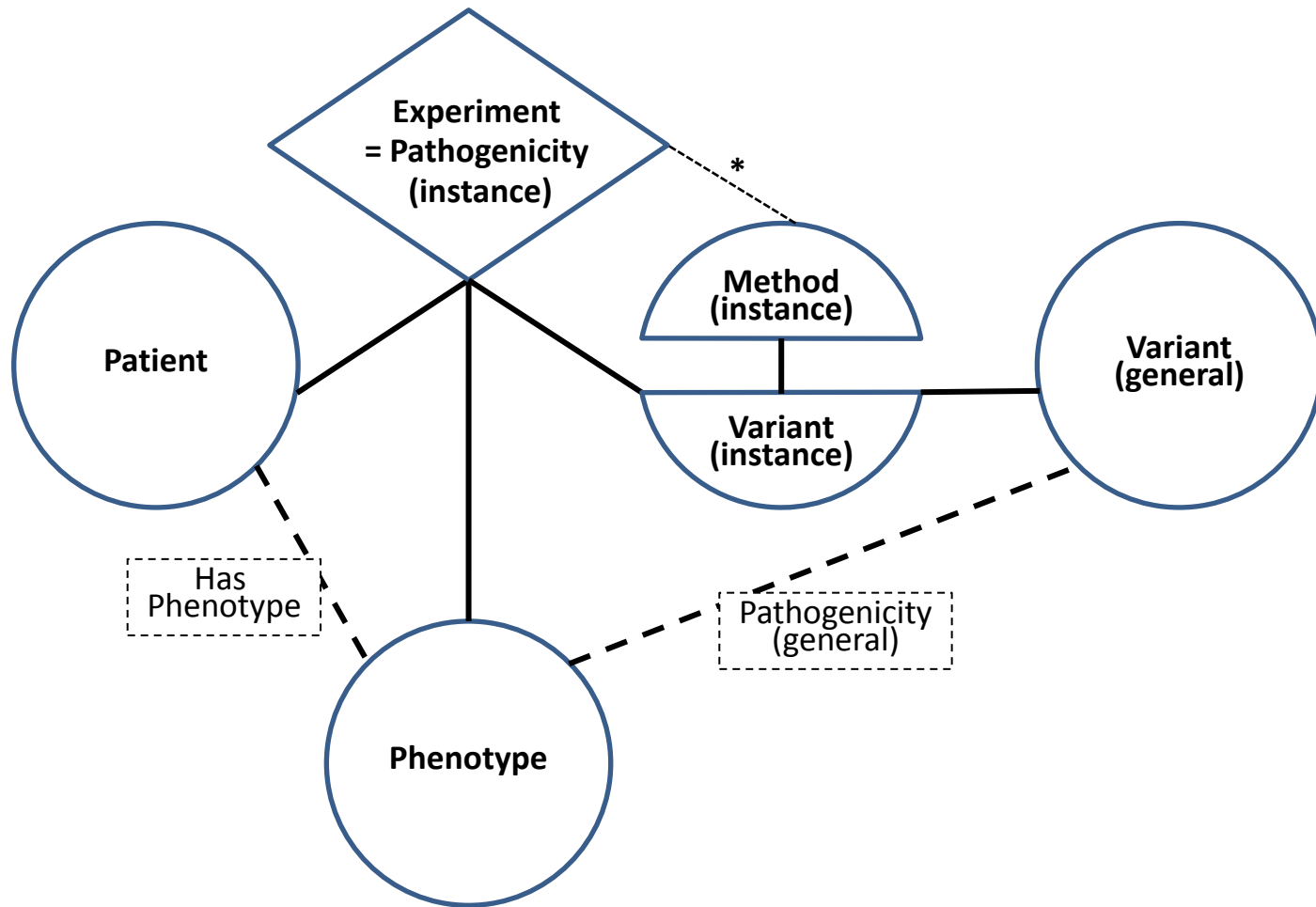


GWAS Central has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 200754 - the GEN2PHEN project.

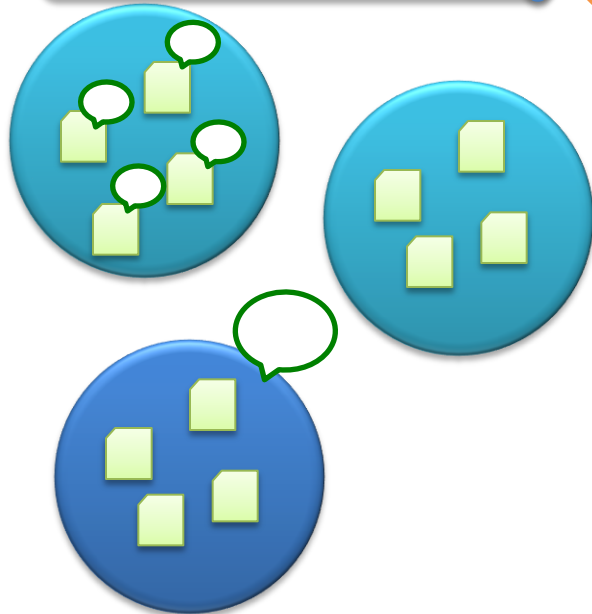


Experiment Centric

Integrates Patient and Variant Centric advantages (and optionally and Method* as well), whilst also providing a place to hold the pathogenicity of the variant in that patient



Search G2P



Feed of search results

Comments and annotations

Resource list



Web services



Web services



Web services

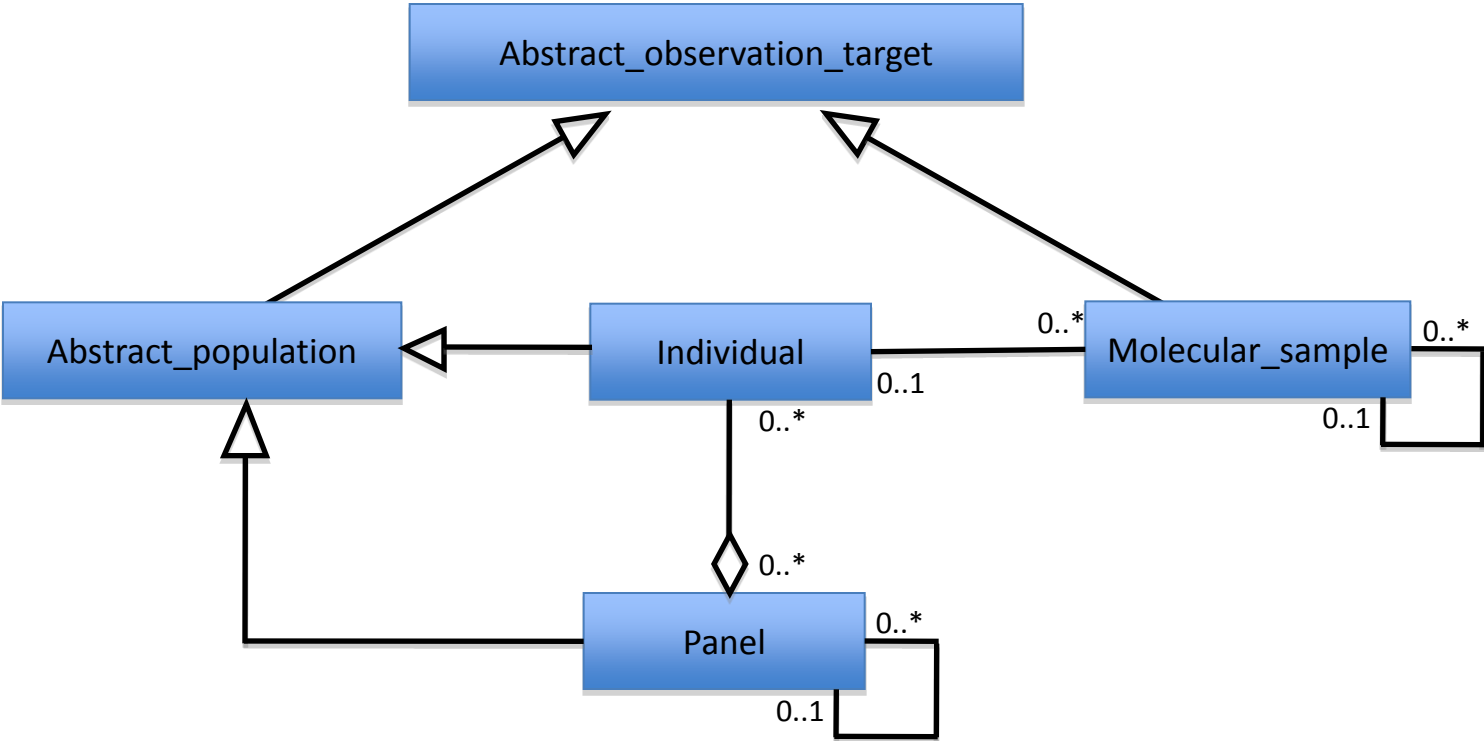


Café Rouge

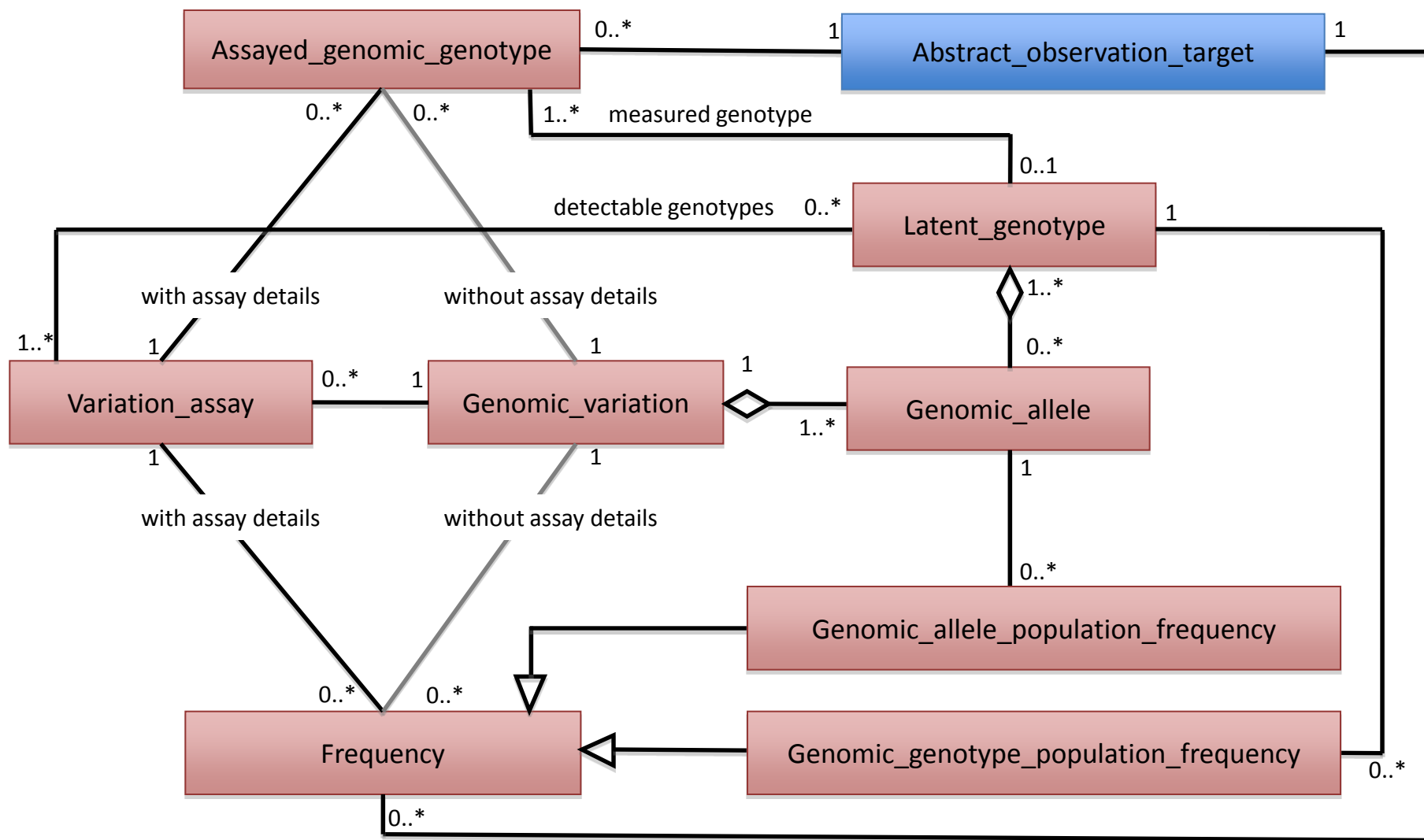


etc.

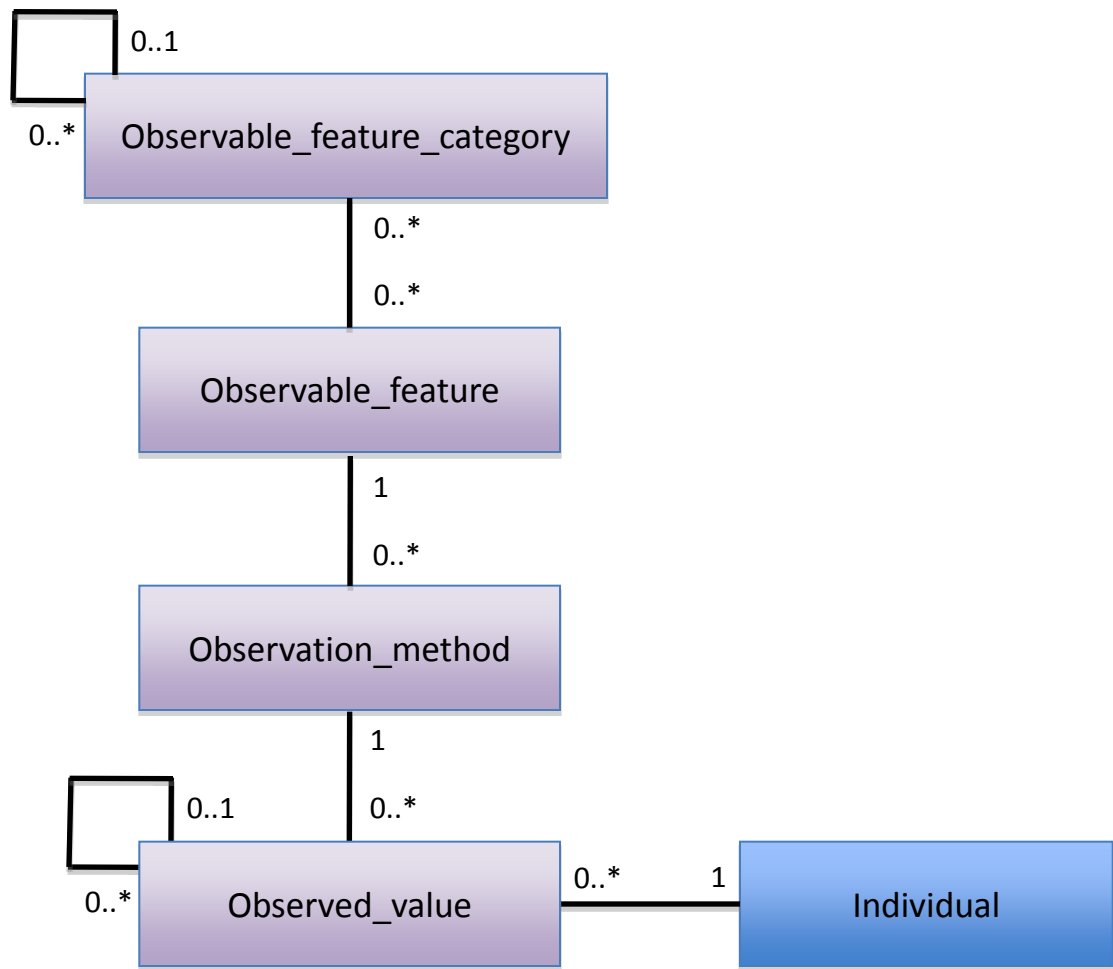
PaGE-OM 'SAMPLE' Domain



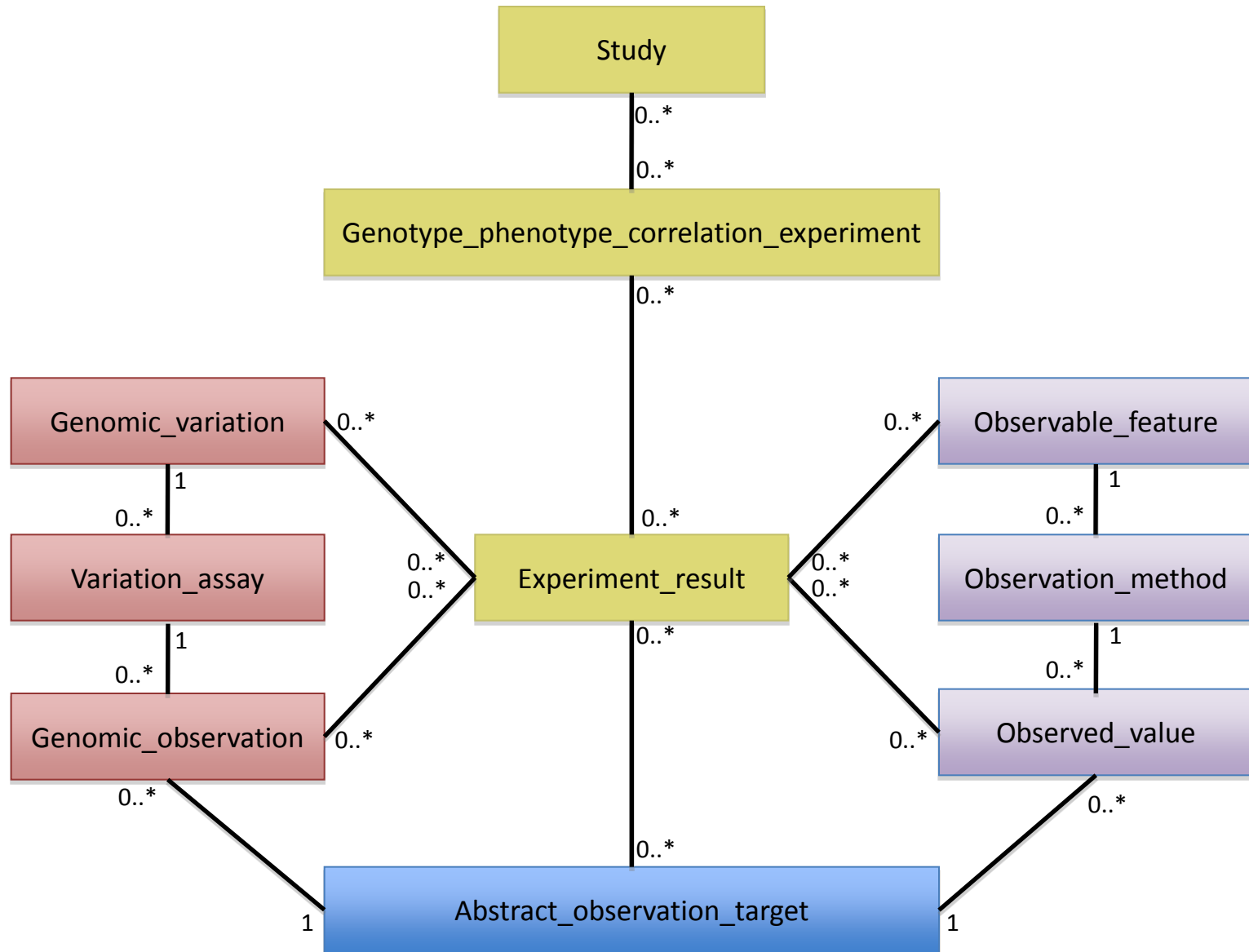
PaGE-OM 'GENOTYPE' Domain

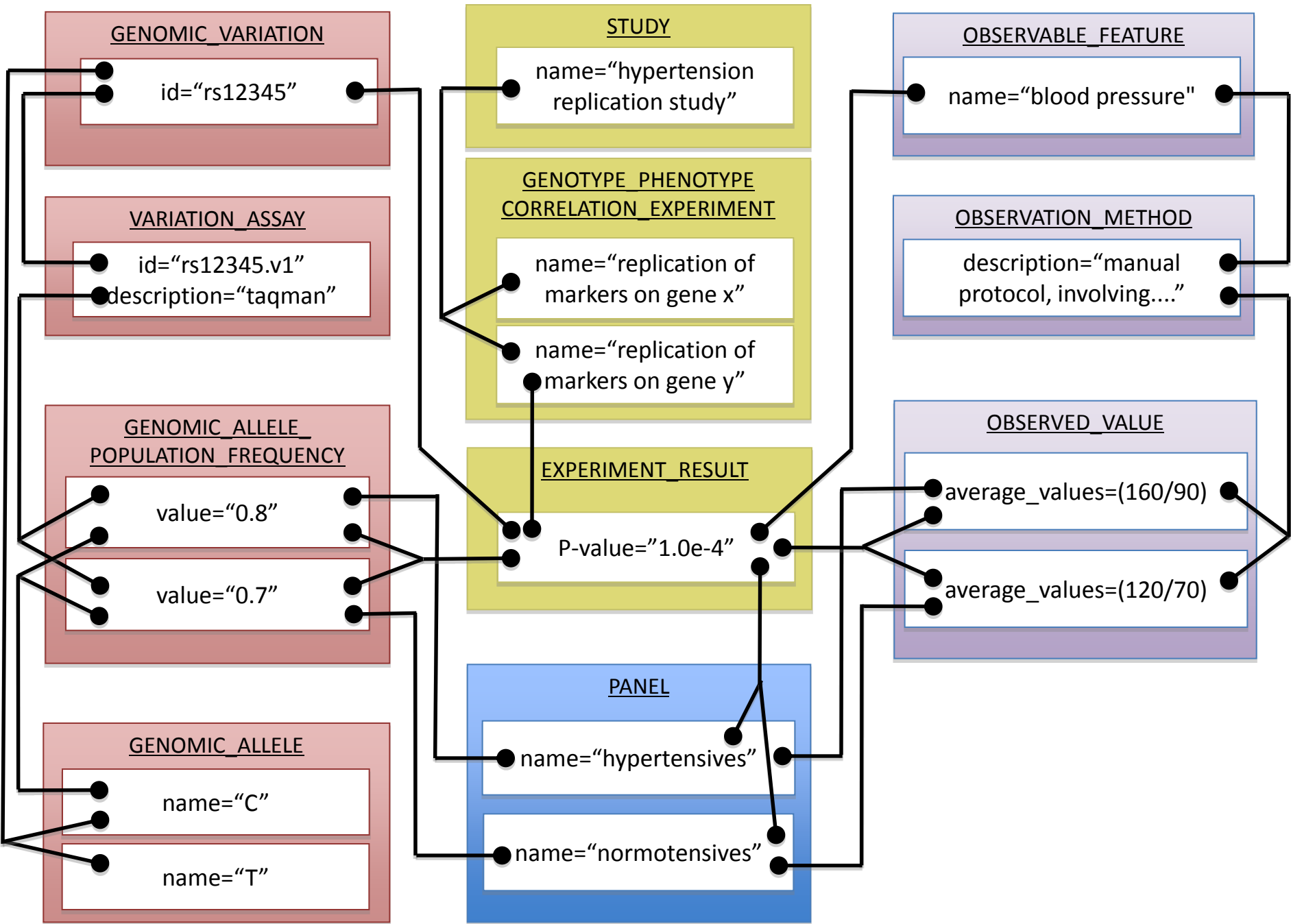


PaGE-OM 'PHENOTYPE' Domain



PaGE-OM 'EXPERIMENT' Domain



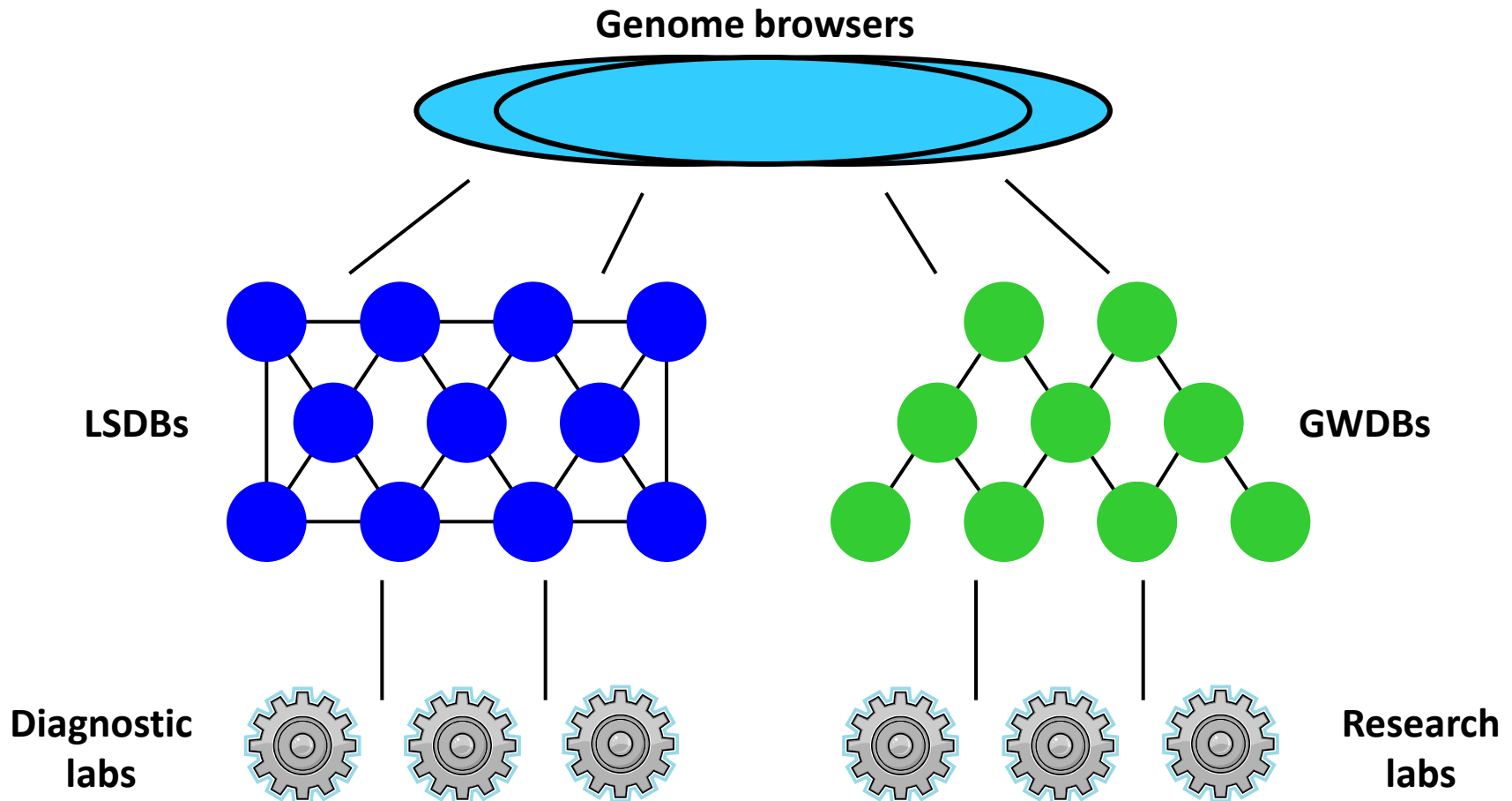


1. Create 'franchised' databases

- data models [e.g. PaGE-OM, Pheno model]
- data management tools [BCP, Phenosys]
- databases [LOVD, UMD, IGVdb, HGVbaseG2P]

2. Build the connections

- ontologies, nomenclatures
- data formats, tools/software
- reference standards [LRG]

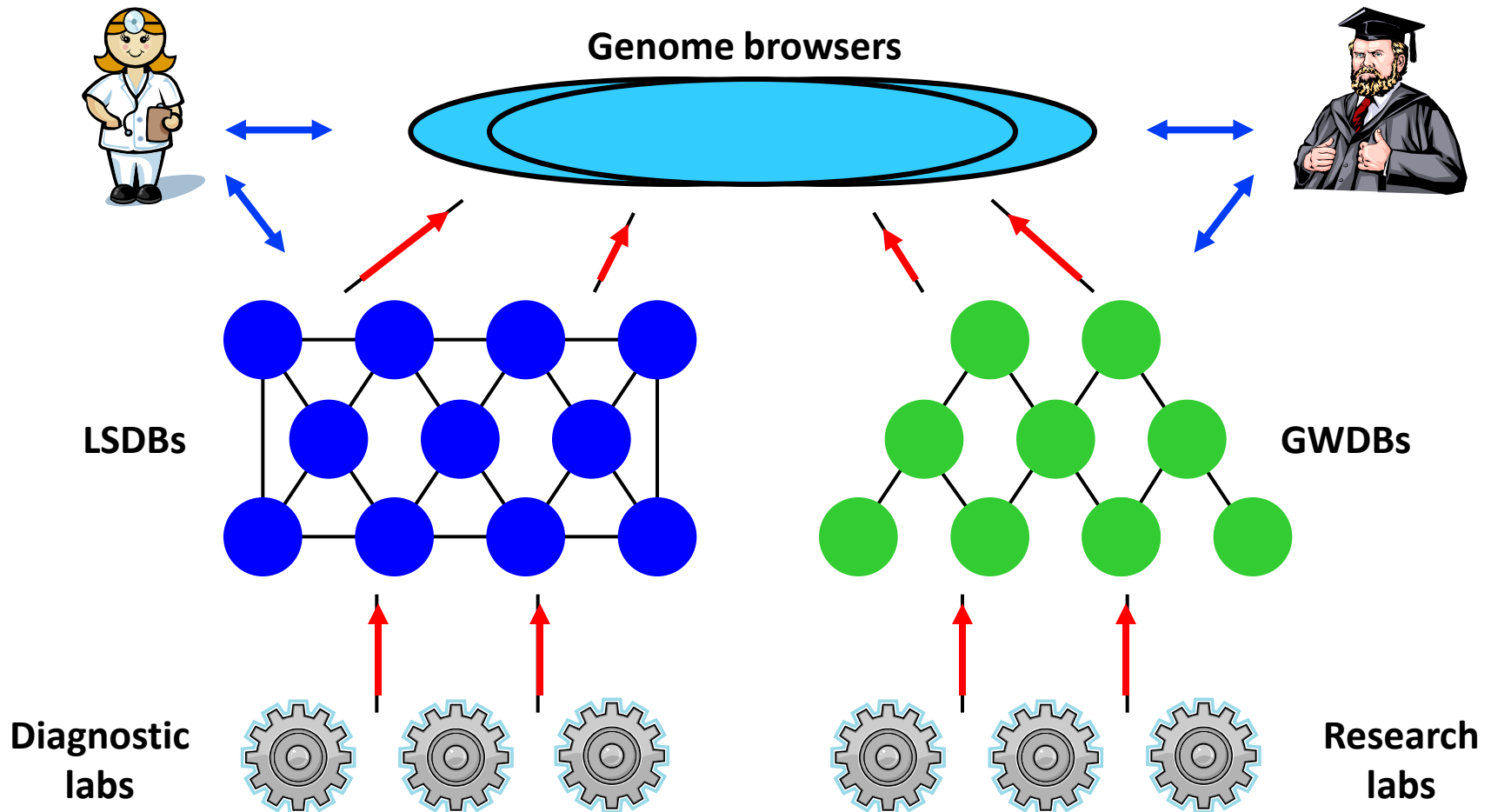


3. Enable the data flow

- legal and ethical [permissions, privacy]
- attribution, incentives, reward [BRIF]

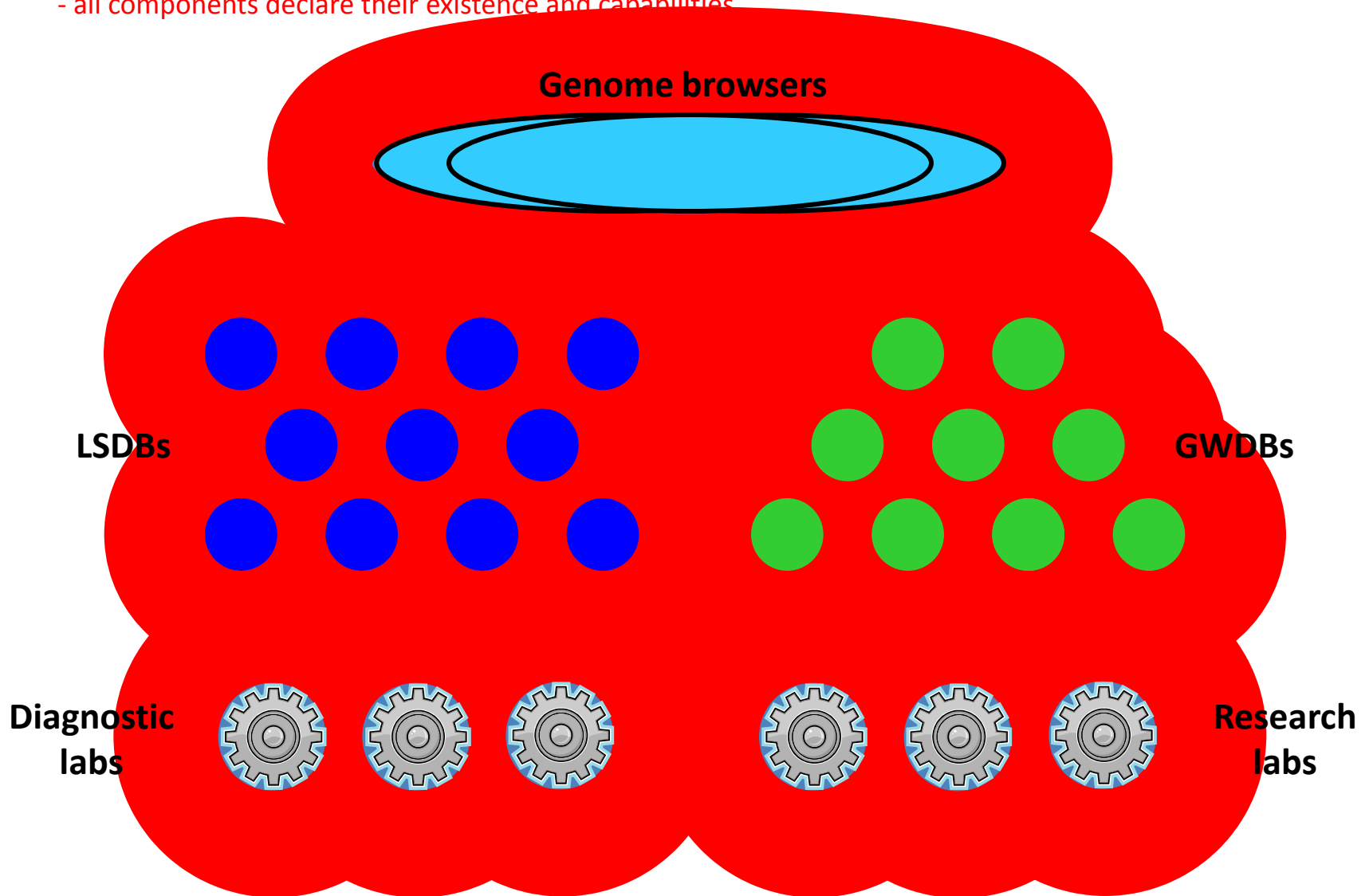
4. Enable data searching

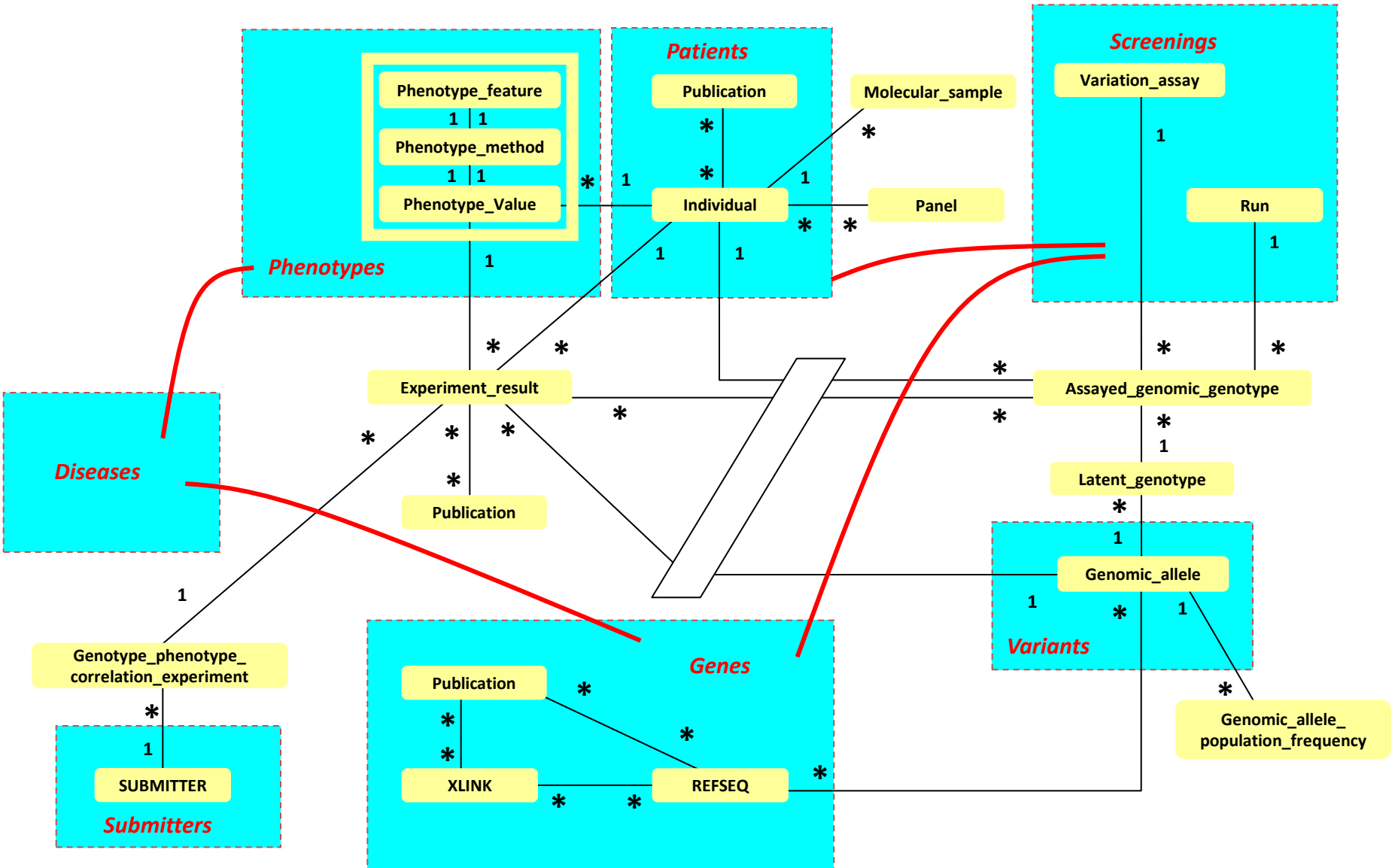
- software [SNP-DAS, APIs, HGVMart]
- interfaces [browsers, DiseaseCard]



5. Grid & semantic web

- workflows, software, security
- permanent global IDs for all 'entities' (people, web pages, pictures, functions...)
- all components declare their existence and capabilities

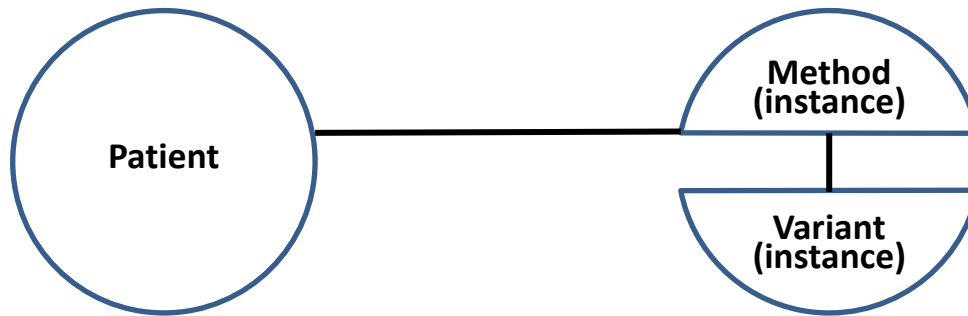


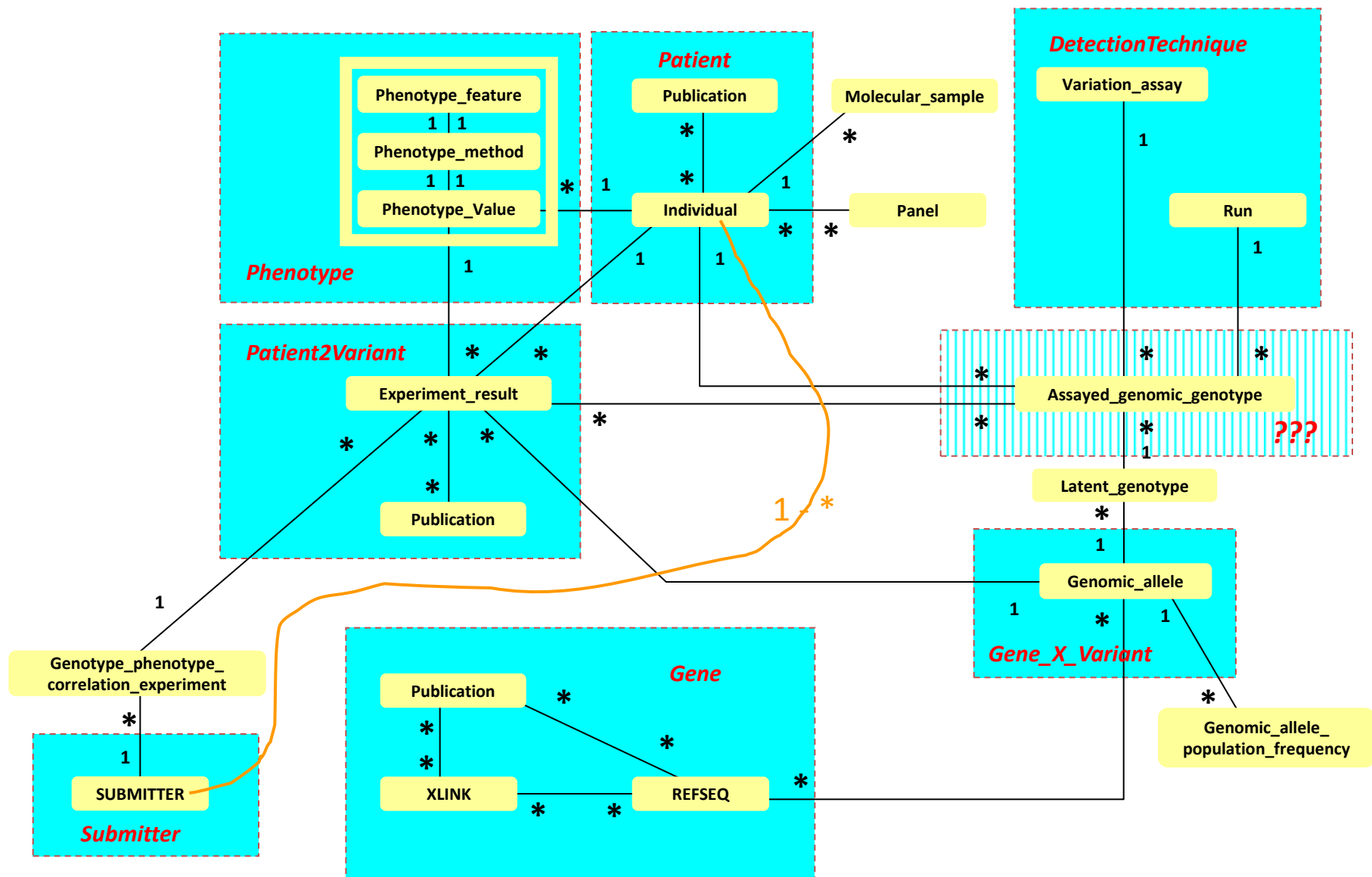


LOVD 3.0

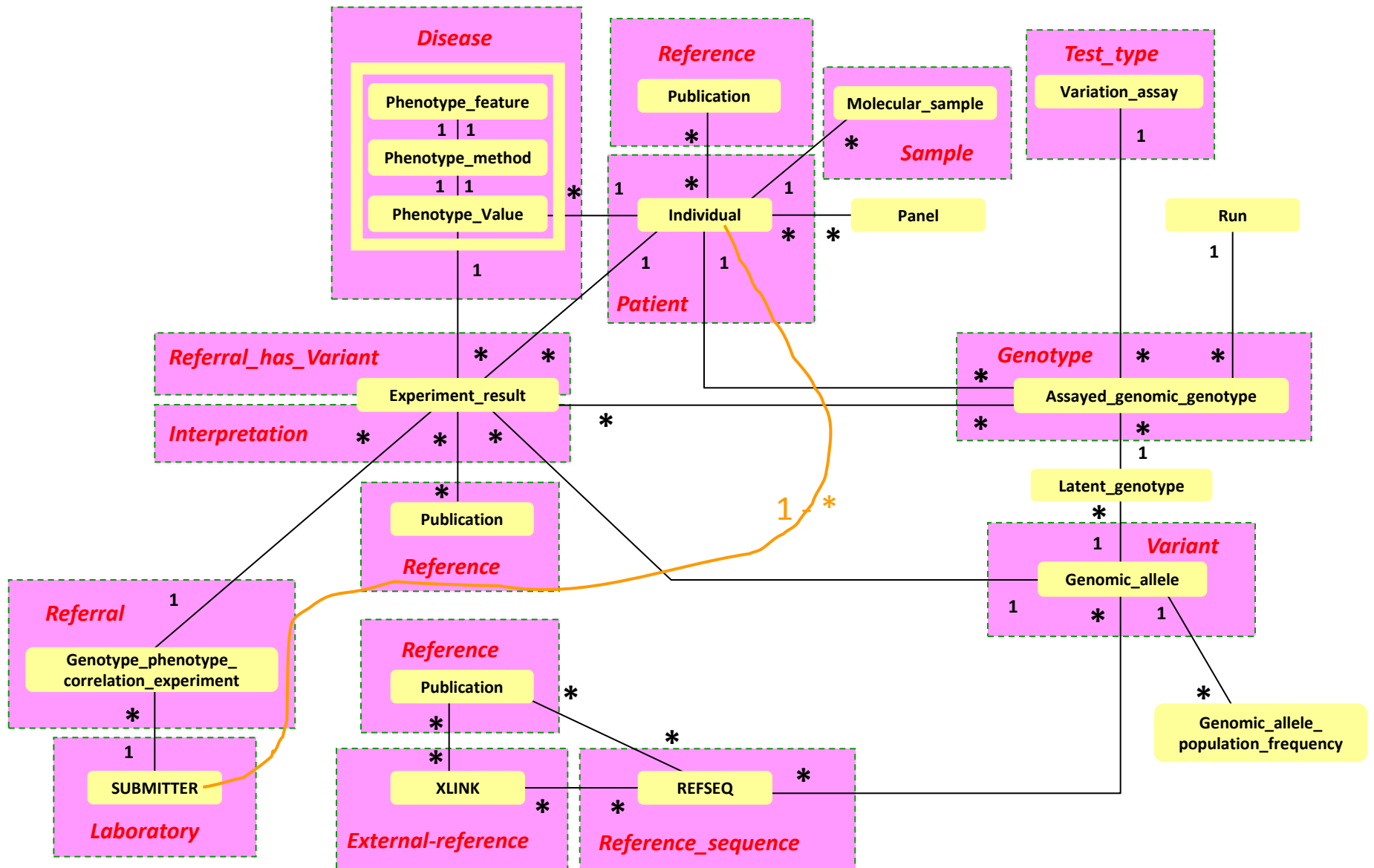
Method Centric (current LOVD 3.0 ?)

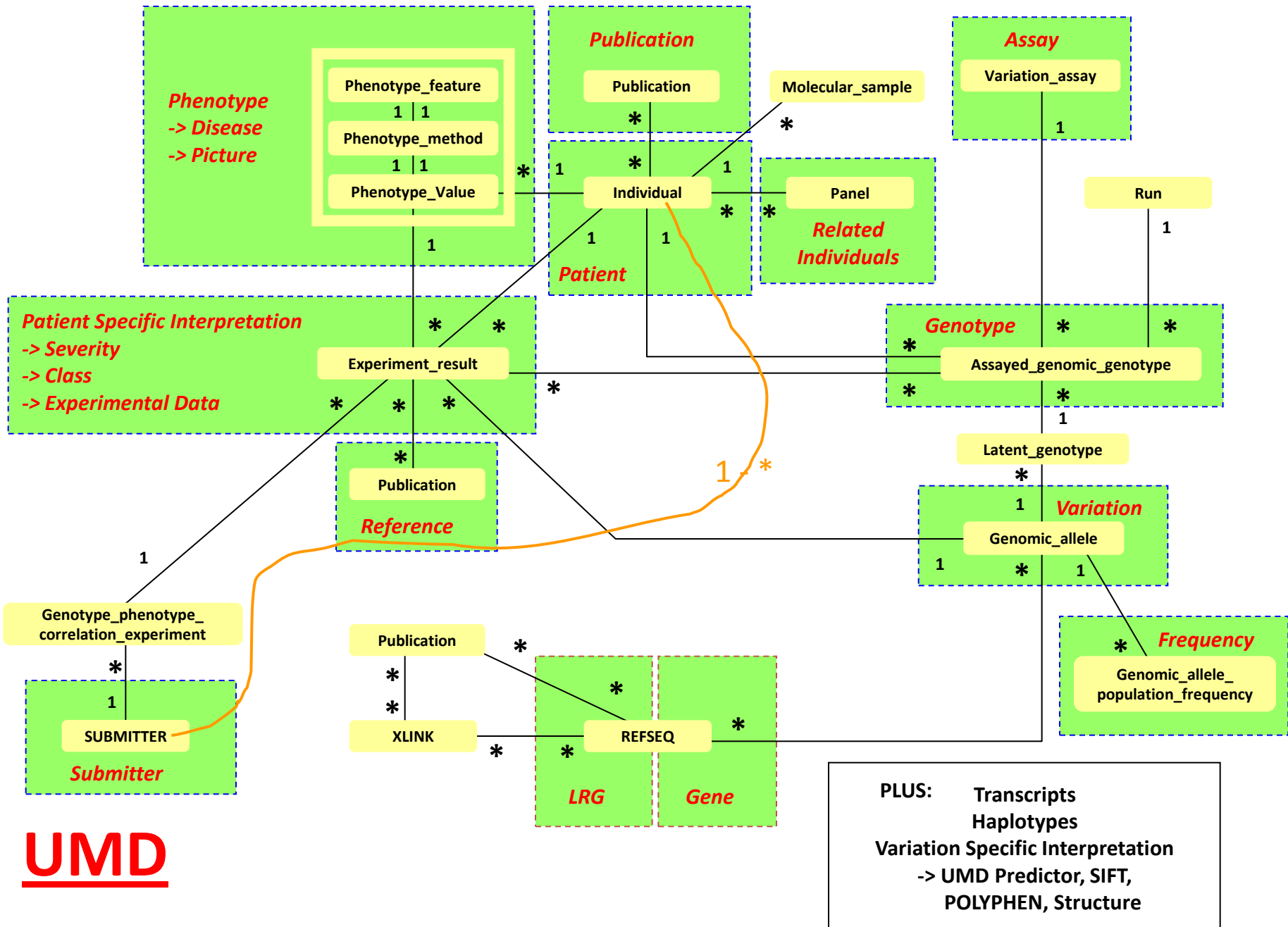
Suitable as a database for labs generating mutation data





LOVD

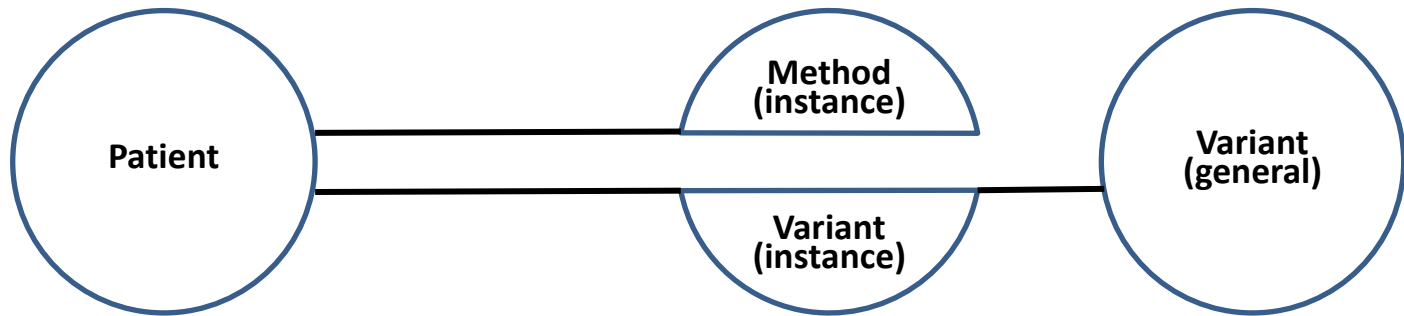






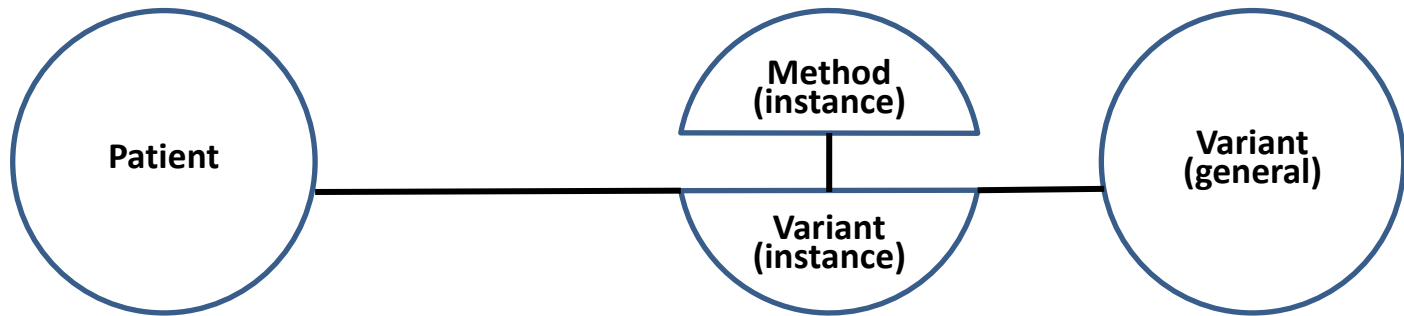
Patient Centric

For the ultimate future, where the genome is sequenced once, and all variants detected



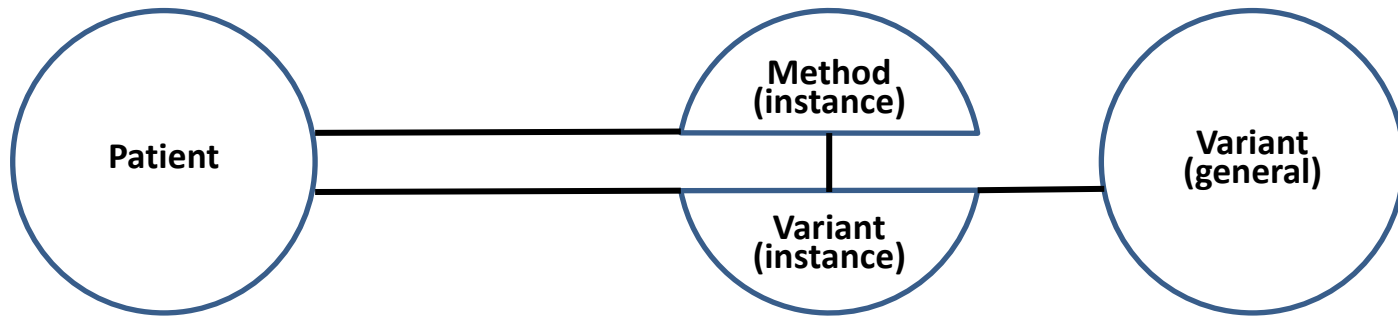
Variant Centric

Old approach, suitable for LSDBs. Can relate to instance or general variants or both



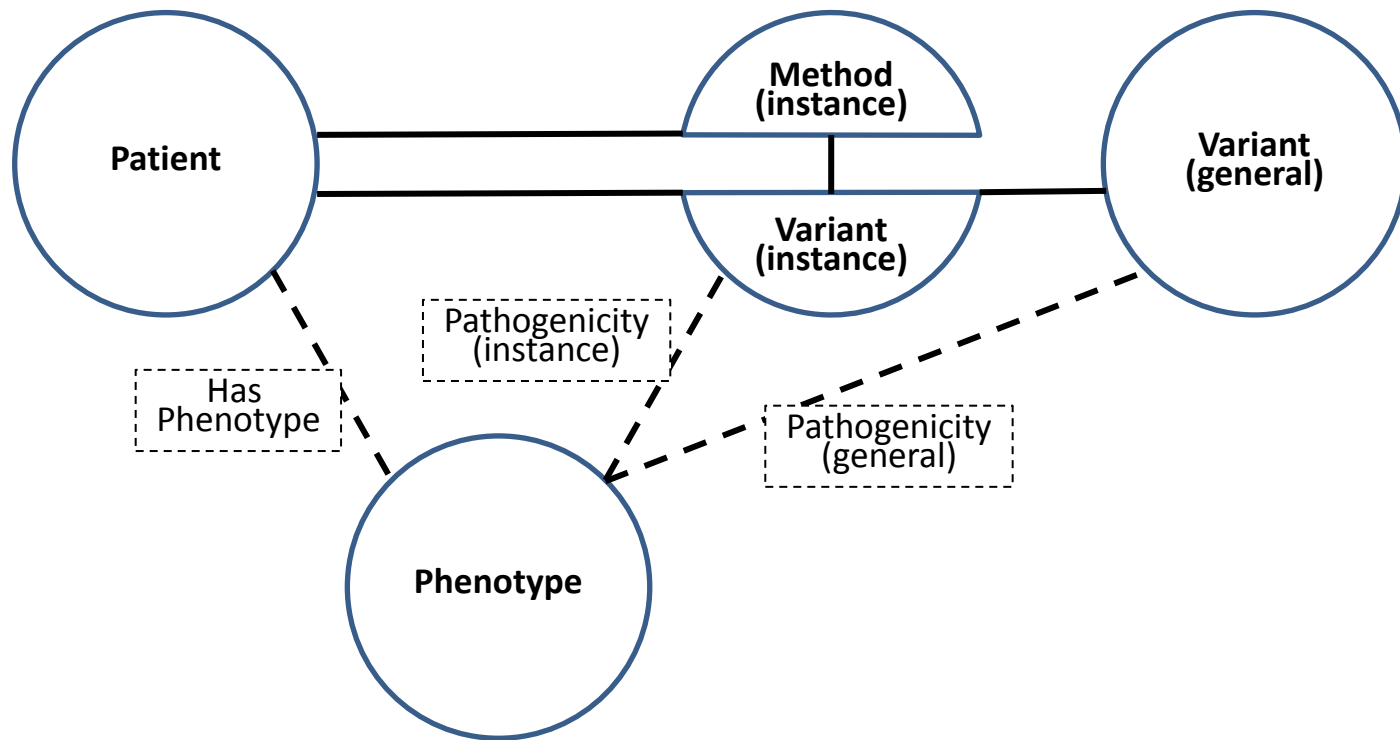
Variant + Patient Centric

Involves redundant relationships, necessarily



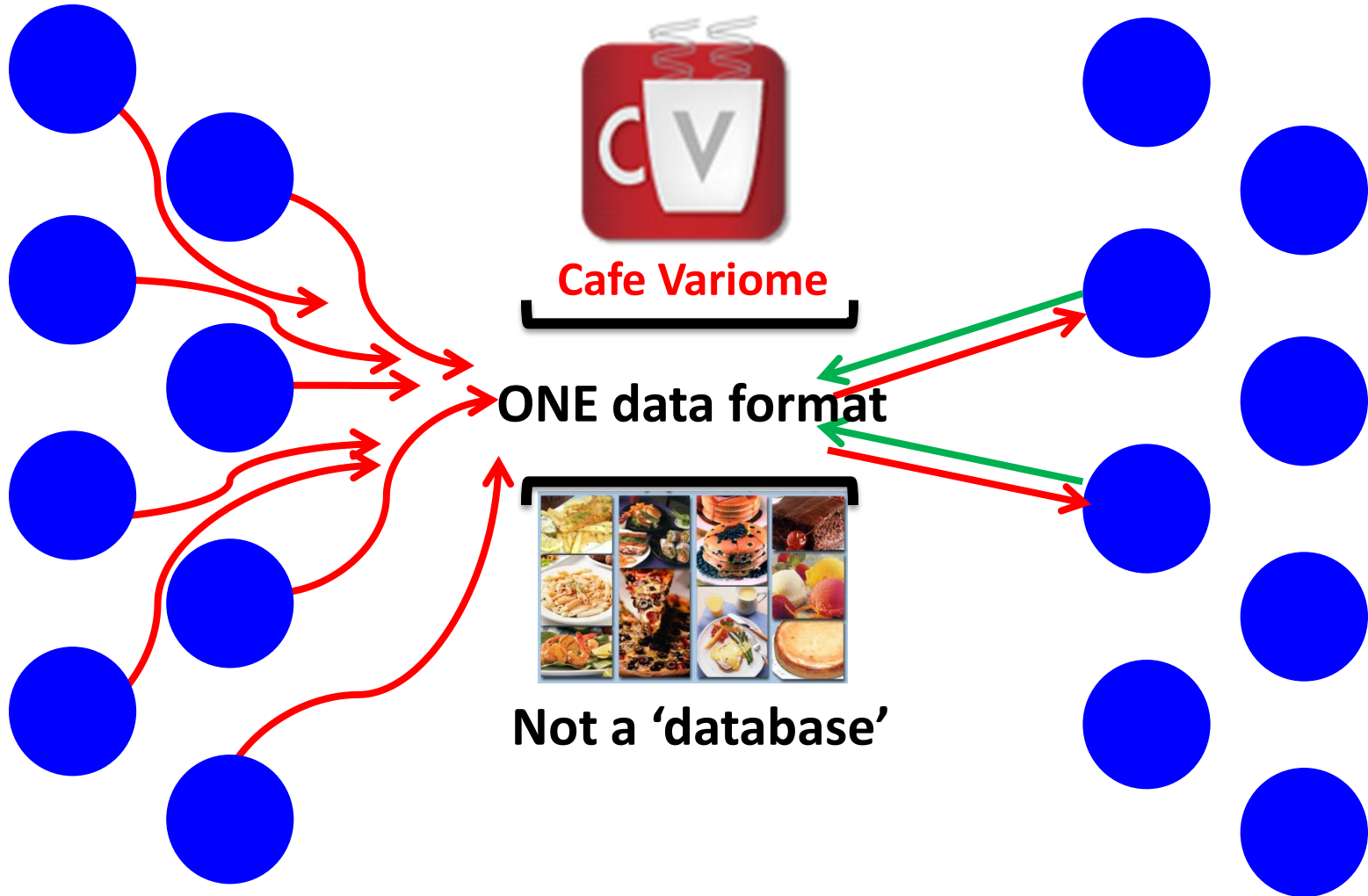
Phenotype Relationships

- 3 objectives, describing:
- a) phenotype of patient,
 - b) variant pathogenicity in patient
 - c) variant pathogenicity in general

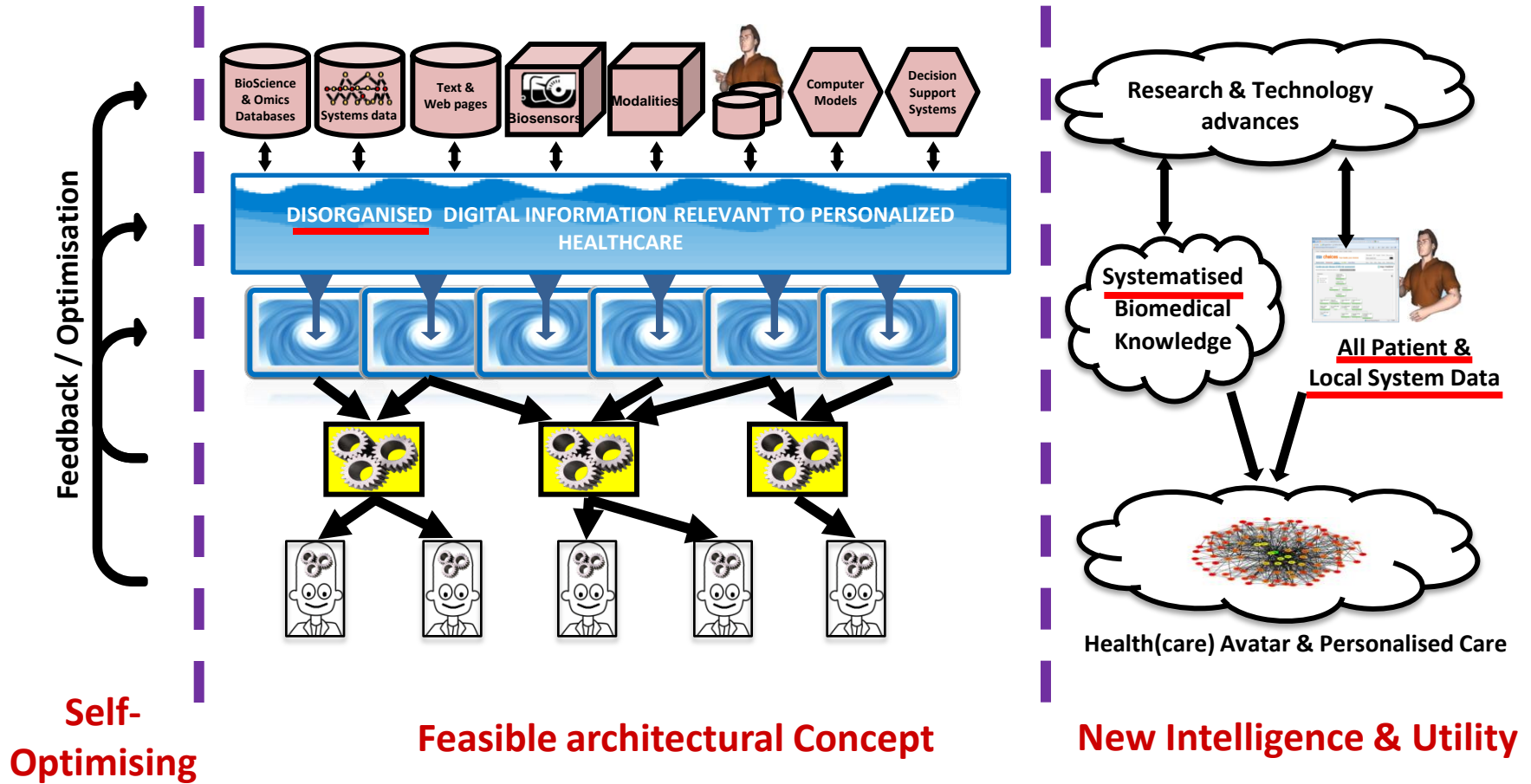


SUBMITTERS

USERS



The I-Health Opportunity



Local &/or Centralised &/or Federated technologies for data display and data mining

Solutions for open sharing:
summary level data, metadata,
& obfuscation strategies

Solutions for controlled sharing:
individual level data,
primary and/or harmonised data

Means for controlled and/or open
data use without sharing:
via DataShield

Eliminate ambiguity, maximise security, and enable recognition/reward:

- Digital IDs for scientific publications (DOIs)
- Digital IDs for Data Releases (DataCite)
- Digital IDs for Researchers (ORCID/OpenID)
- Digital IDs for BioResources (BRIF)

Tool for discovery of
sample collections + original + harmonised variables + counts/means

DataShaper
development and use

Web services

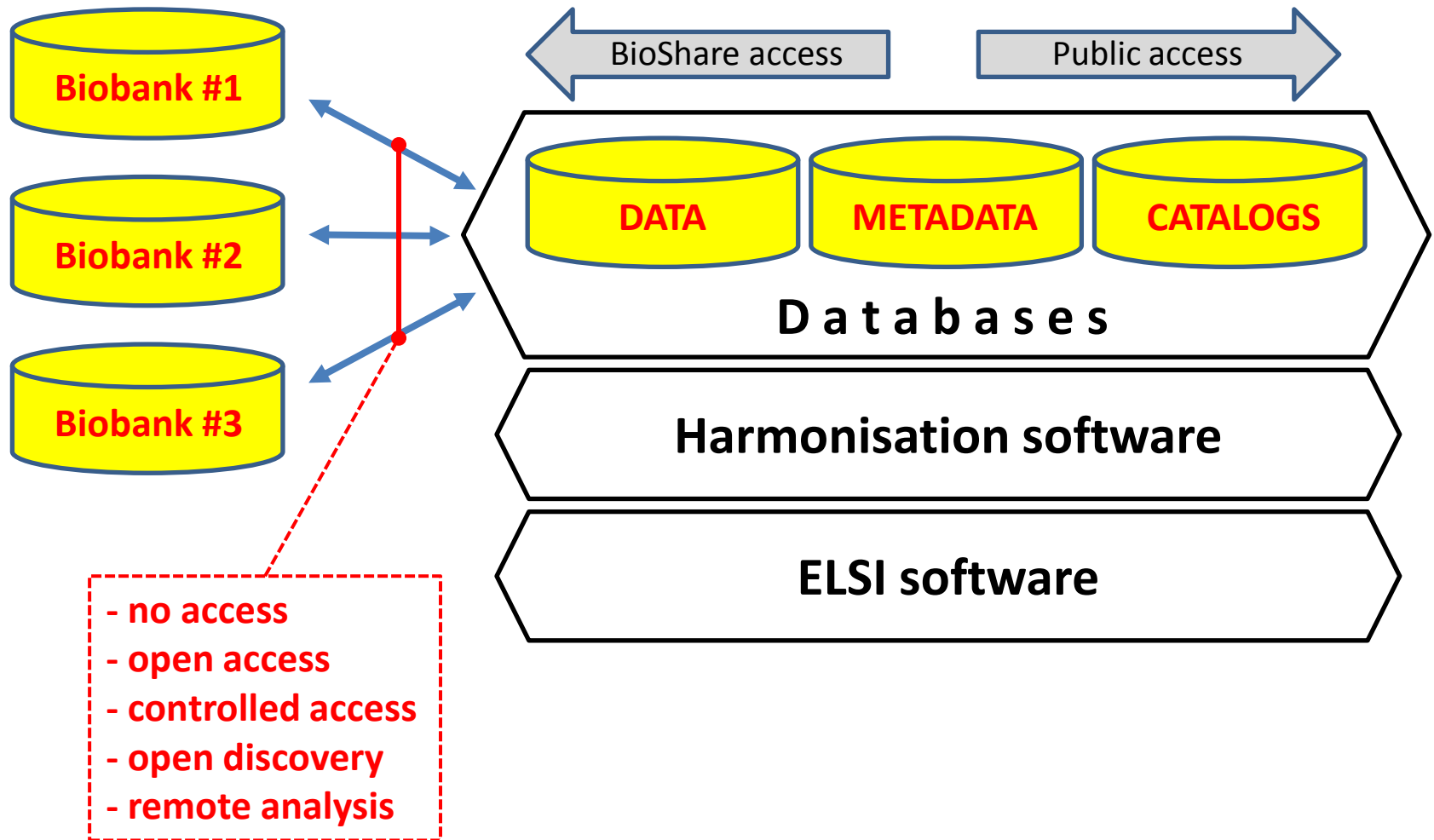
New database for
sample collections,
variables + results

Web services

Existing database for
sample collections,
variables + results

Web services

Existing database for
sample collections,
variables + results



Need: Digital 'Big-picture' across diseases/services/self-care/pathways

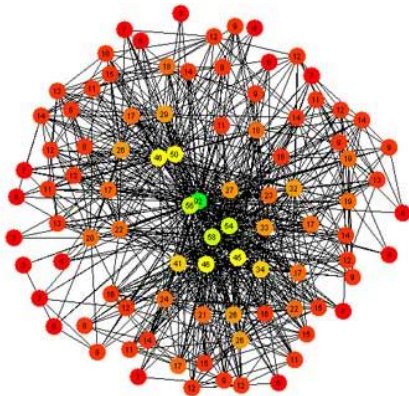
Diabetology:
Glucose control

Ophthalmology:
Diabetic eye care

Nephrology:
Chronic kidney disease

Key research knowledge

Patient Biometrics



Future: Realistically complex and
dynamic model/avatar of "Mr Smith"

RESEARCH DATA

Omics data
Systems studies
Computer models
Biobanks/Registries
Clinical trials
Disease research
Drug research
Epidemiology
Animal models

HEALTHCARE DATA

EHR content
Medical publications
Medical websites / blogs
Protocols / guidelines
Diagnostic test results
Biosensors outputs
Lifestyle data
Environment data
Drug /treatment info

DIGITAL INFORMATION RELEVANT TO PERSONALIZED HEALTHCARE

ICT
'gap'

RESEARCH USE

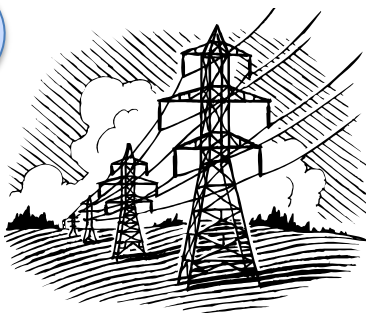
HEALTHCARE USE

I-Health Challenge: Three clouds

*...bring together **people, methods, and research + patient data**
across molecular, clinical and population scales*

People with
relevant expertise
and authorisation

Intelligence



State-of-the-art
algorithms

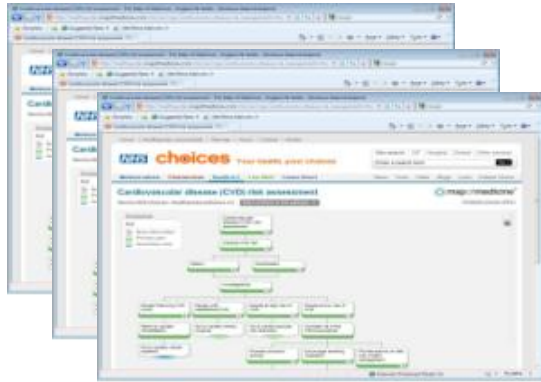
Quality assured
integrated data

Data-2-Knowledge-2-Practice Centre

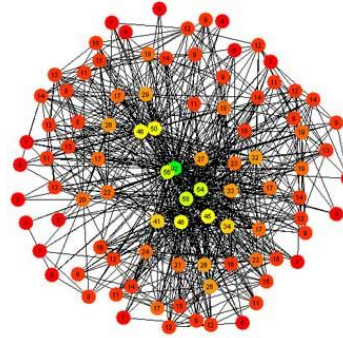
Two floors of biobank & I-Health IT, atop a CVD & respiratory disease clinic PLUS advanced biobank



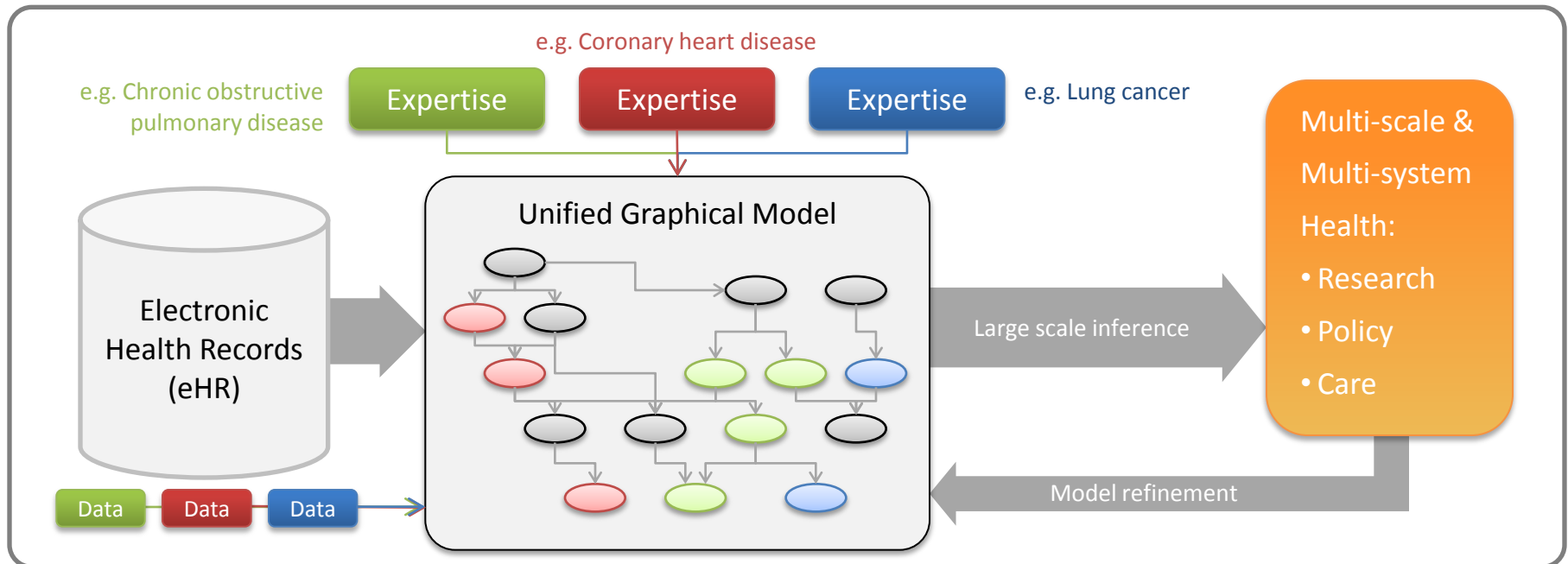
Health Records & Knowledge Silos



Open Unifying Modelling: Across mechanisms and contexts



Health Avatars & Dynamic Models



Central DBs

Federated DBs

