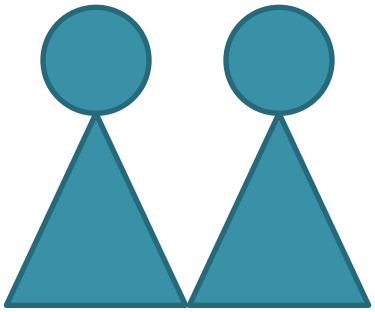
Initiatives to Speed up Data Sharing

- Anthony Brookes
- ICORD 2014, Ede, Netherlands
- 9 October 2014

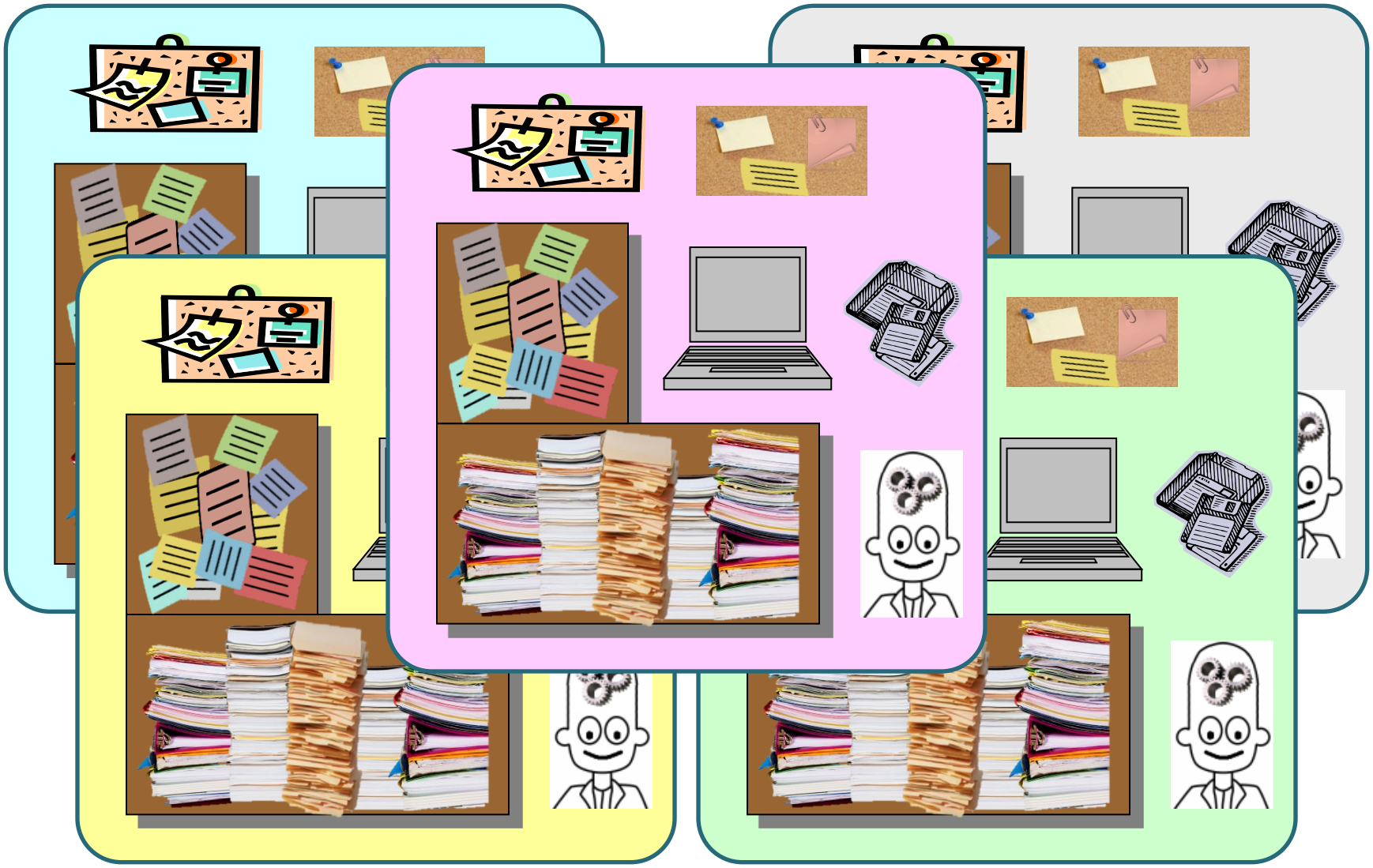
IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



Single Team

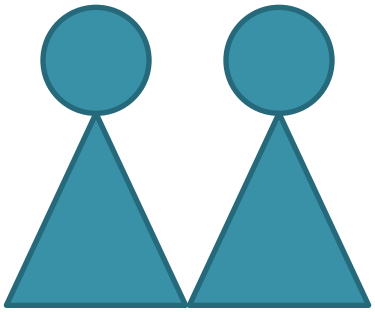
Data

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

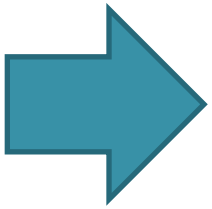


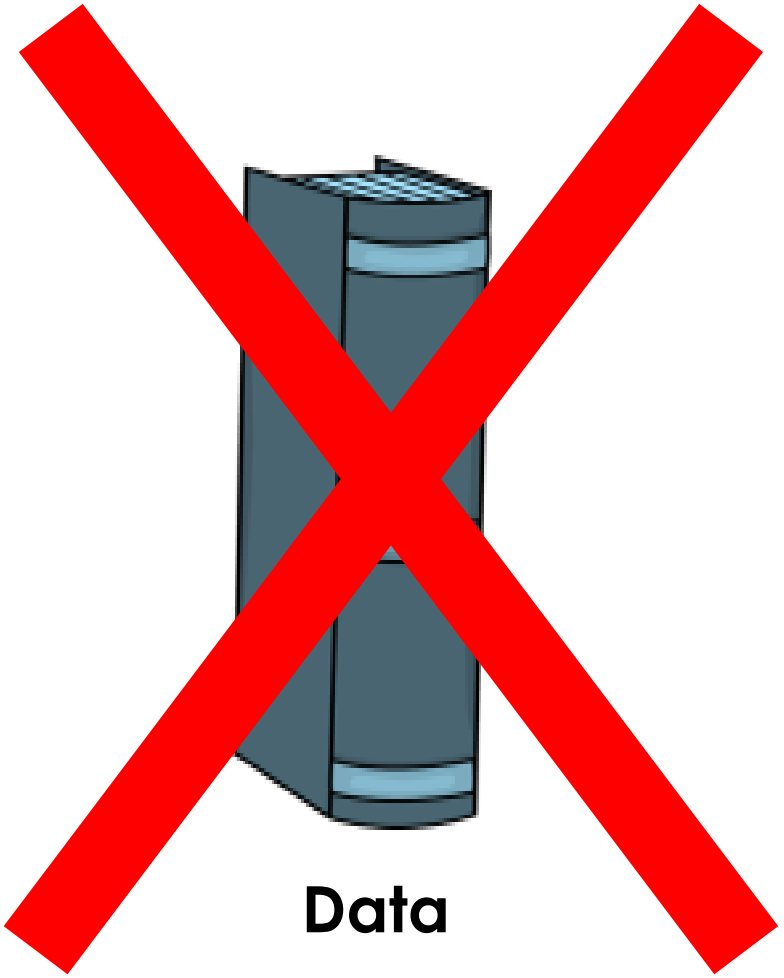
Consortia / Networks

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



Single Team





Data

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Federated

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Rationale

- Research & healthcare produces an enormous amount of data
- These data have significant ‘value’ – unless shared we cannot develop optimum knowledge, diagnostics and treatments, while ensuring efficient utilization of scarce resources
- Datasets include data and metadata relating to phenotypes, genomic variants, other ‘omic’ data, natural histories, clinical trial data, etc
- Resources include patient and family material (extracted DNA, cell lines, pathological samples), technical protocols, informatics infrastructure, and analysis tools

Barriers to Data Sharing

- Technical and Fi issues
  - Storing v ering, securing terabytes
  - Providi gistics for sharing data
- Ethical and Legal is
  - Absent/Uncl haring
  - Bridg ivate networks
  - Differe onal rules/policies
- Cultural issues
  - Reluctan valuable asset (resea cians/institutions)

CAN NOT

MUST NOT

WILL NOT

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

Openly Shared

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

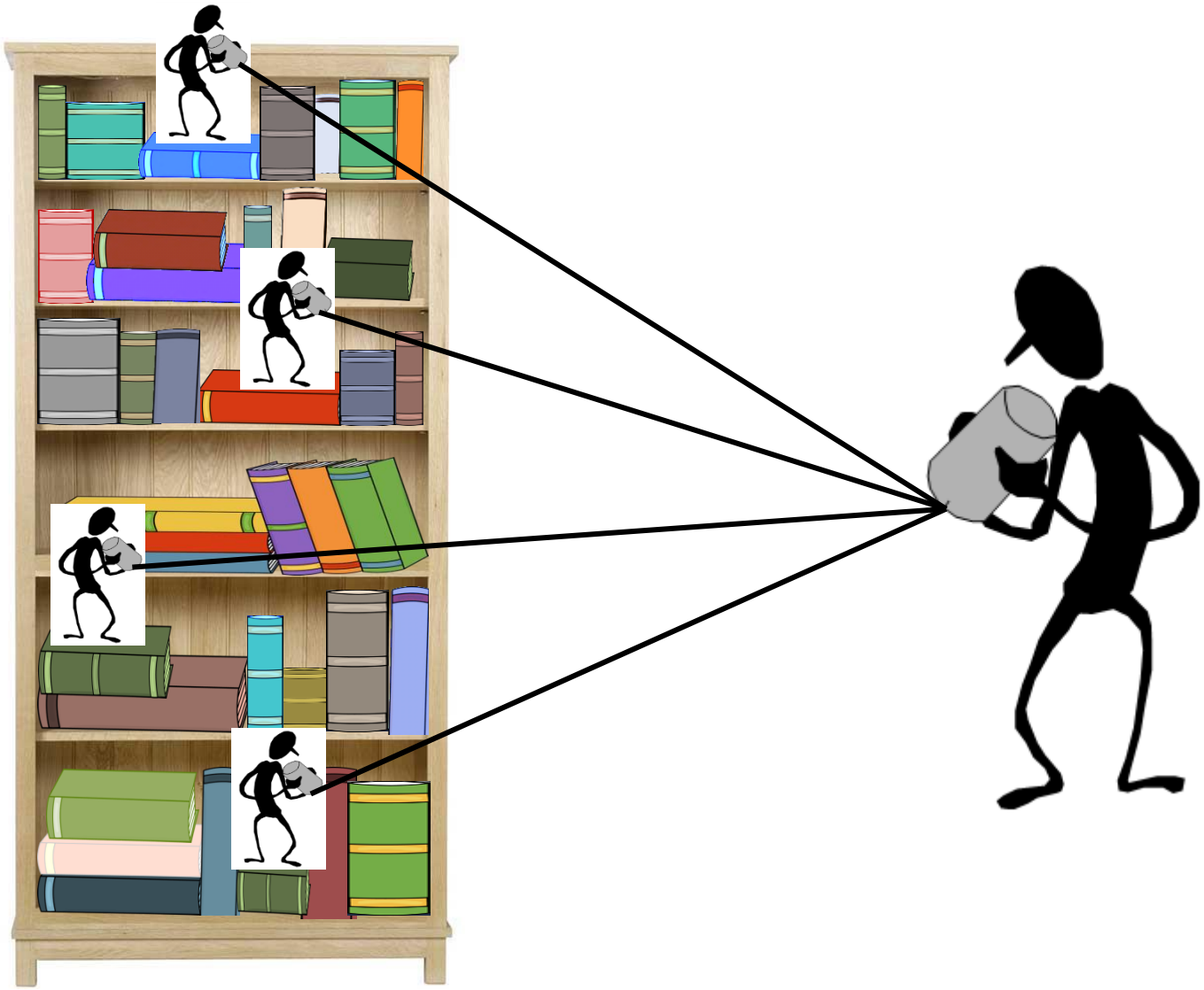
IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Controlled Access

Open Access and Controlled Access approaches are important....

....but other options exist, which are not only useful but arguably essential

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



Remote Data Analysis

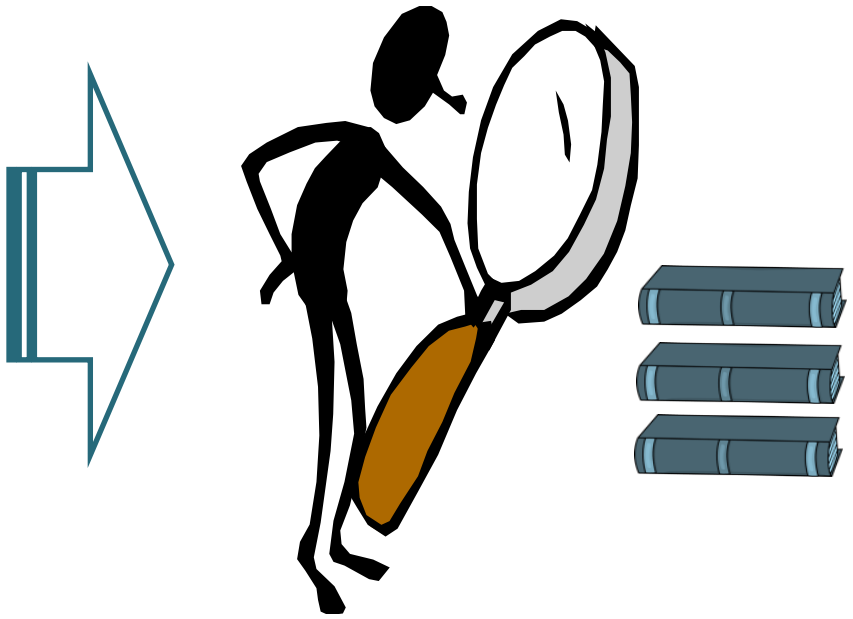
IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Focus on Knowledge & Visualisation

Find the Data (catalog approaches)

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



Data Discovery

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

IRDiRC principles (1)

Sharing and collaborative work

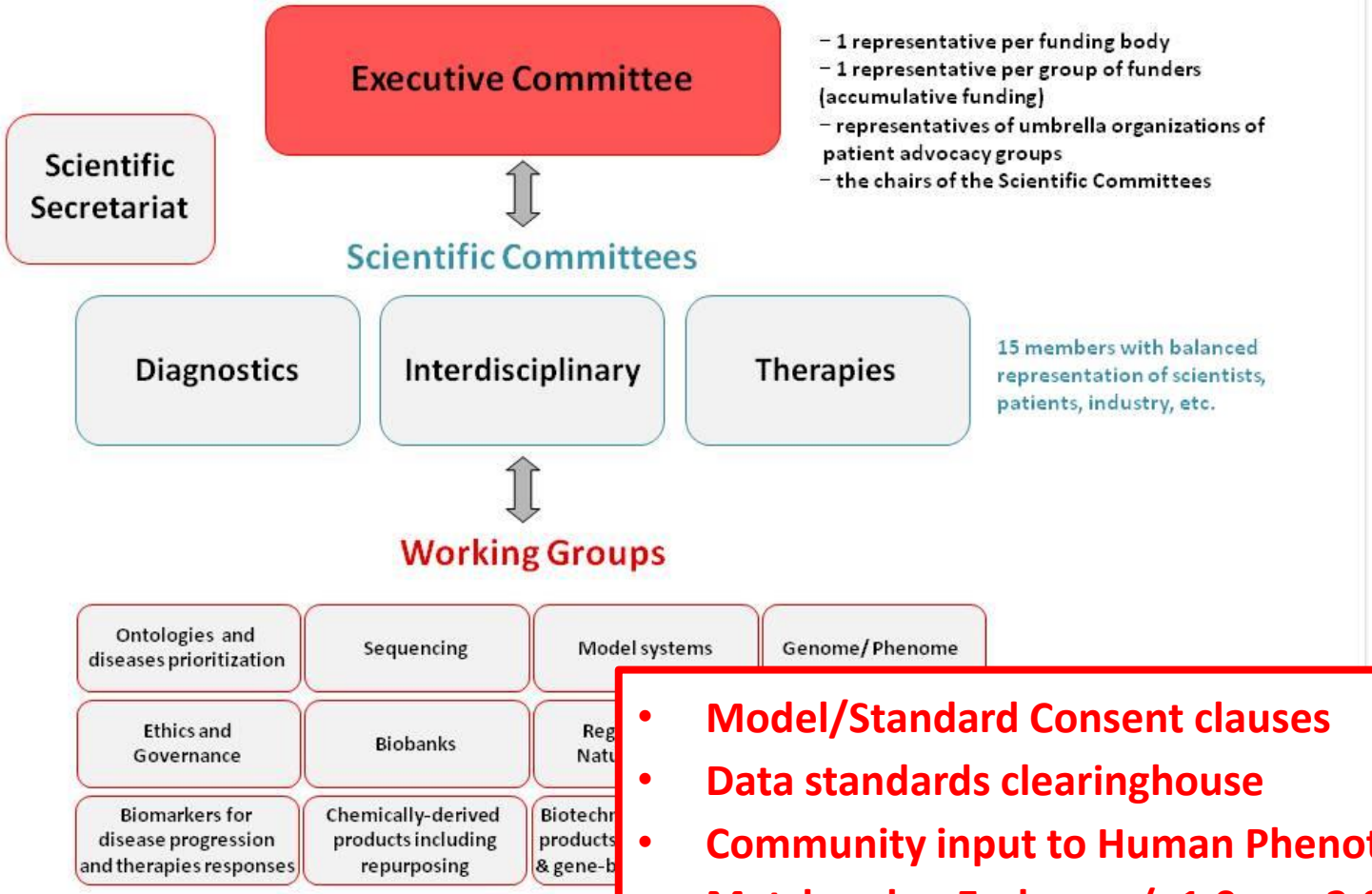
- Informed consent based procedures
- Rapid release of data
- Interoperability of data
- Using maximally open access databases

Scientific standards, requirements and regulations

- Projects should adhere to existing standards
- Standard development - especially consents, metadata, minimal content, patient IDs, ontologies & validated biomarkers
- Upon sharing/publication, always cite utilised databases & biobanks

IRDiRC principles (2)

- **Guidelines & Policies:**
  - Data generated from research projects, including source data, should be deposited in appropriate open or controlled access public databases.
  - Projects should adhere to standards endorsed by IRDiRC
  - Data producers acknowledge their responsibilities to release data rapidly and to publish initial analyses in a timely manner.



- **Model/Standard Consent clauses**
- **Data standards clearinghouse**
- **Community input to Human Phenotype Ontology**
- **Matchmaker Exchange (v1.0 => v2.0)**
- **Minimum open data conventions**
- **Globally unique patient IDs**
- **Bio-resource metrics and impact system**
- **Population Controls Variant Resource**

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

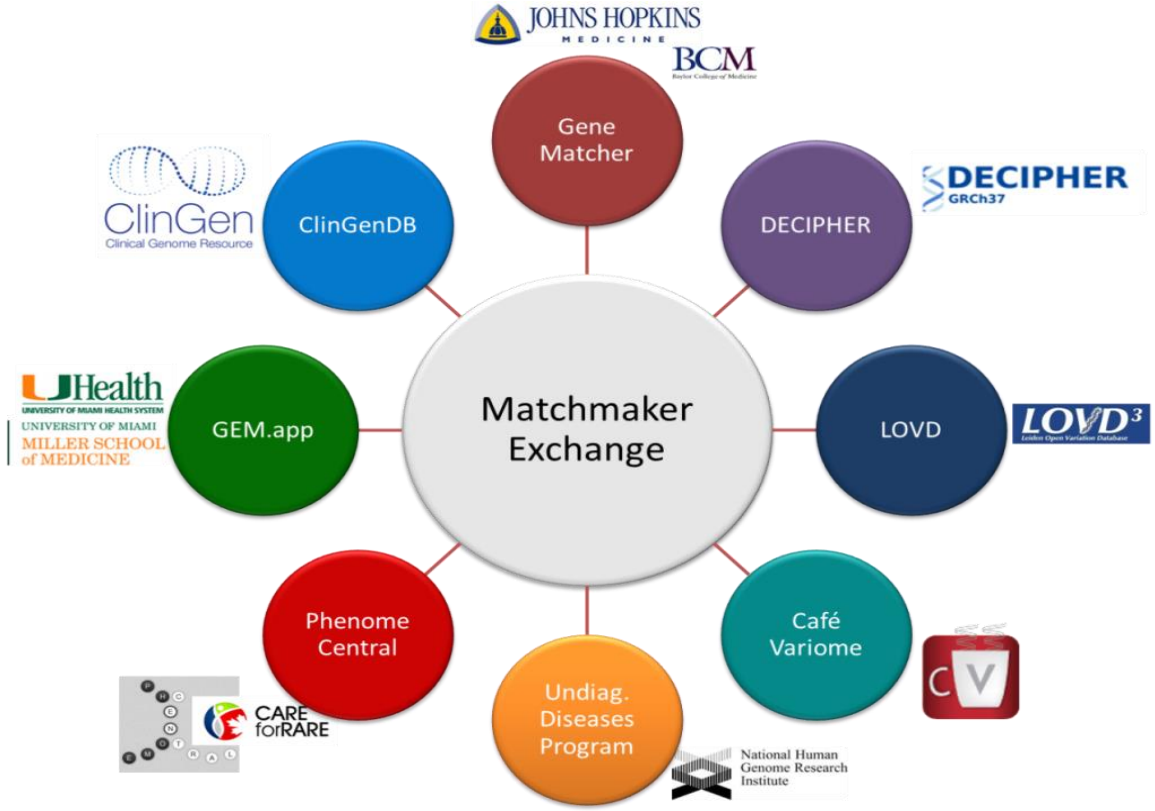
ClearingHouse project (1)

- Catalog: list of available data standards
  - In development – release in January 2015 on the IRDiRC website
- Dynamic: catalog with an added dimension
  - Contains a list of data standards currently available
  - Allows users to select the most appropriate data standards to apply
  - Allows standards submissions from users
  - Allows users to connect/collaborate on standards creation
  - Includes use cases and exemplar applications
  - Informs researchers about standards outside their specific RD fields

ClearingHouse project (2)

- Five main fields of application
  - Standards in Genomics and other OMICS
  - Standards in Phenotyping
  - Standards in Outcome Measures for clinical trials
  - Standards in Human Data Registration
  - Open and controlled access databases to store data
- Align with other ongoing efforts
  - RD-Connect, Orphanet
  - PCORI, Comete
  - ELIXIR, BBMRI, GA4GH, Data FAIRport

Matchmaker Exchange (1)



In research and clinical settings, many RD patients are difficult to diagnose in isolation. Finding a similar (geno and/or pheno) patient elsewhere may provide sufficient evidence to identify the causative gene. 'Matchmaker Exchange' aims to enable those similar patient pairings to be identified via a standardized APIs and procedural conventions.

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

Matchmaker Exchange (2)

- **Tiered Informed Consent**
  - Explicit patient consent may not be needed, dependent on the potentially identifiable information made discoverable
  - Subsequent sharing would need explicit informed consent
- **Requirements & conventions**
  - Data provider and match requestor MUST provide contact details
  - Recommended that searches and matches are logged
- **'Data Submission' API**
  - Requestor sends in their patient's data, these data are retained for future matches, basis of match is not controllable or revealed
- **'Data Query' API**
  - Requestor sends in a question and thereby controls the basis of the match, with an option to deposit data to enable future matches

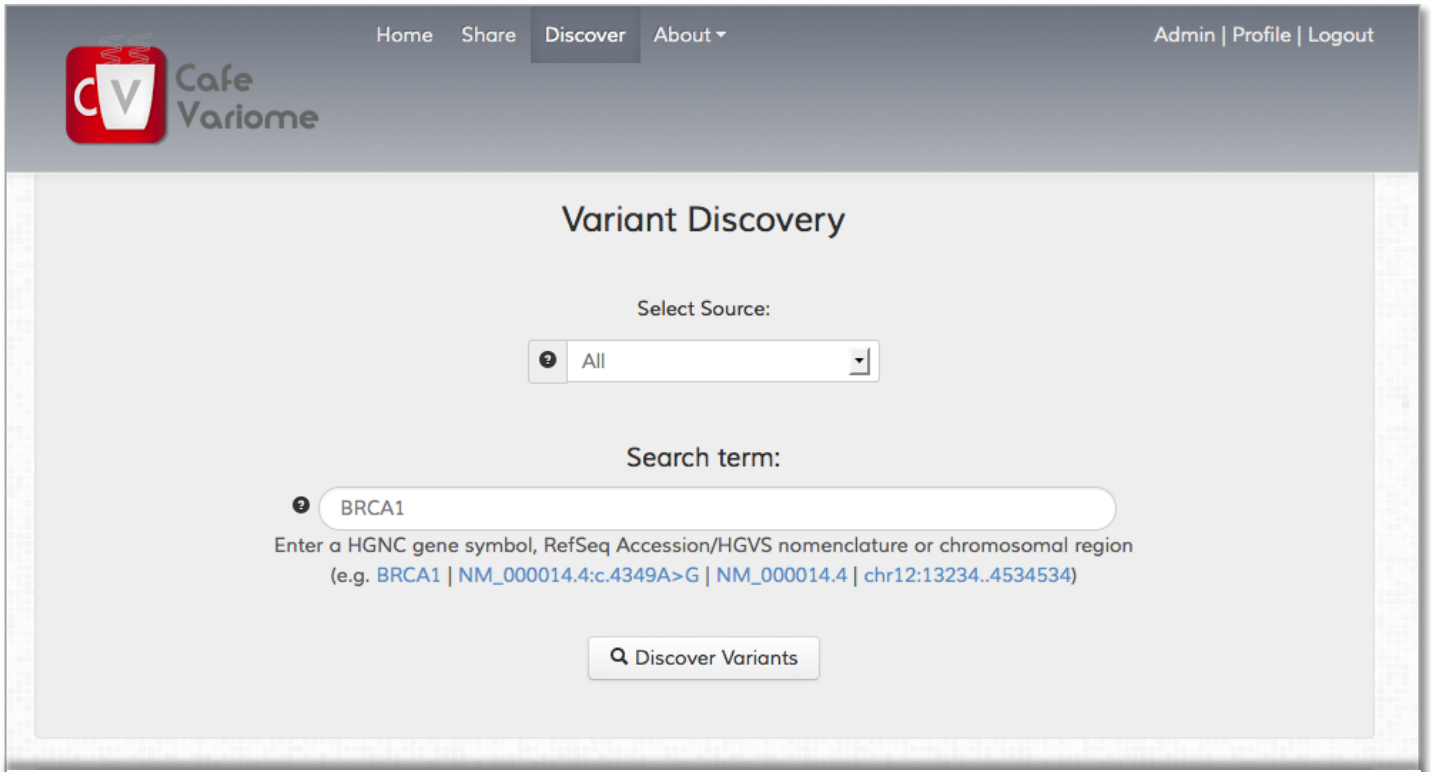
Café Variome (1)

Share the 'existence' rather than the 'substance' of data

This technology sits atop/alongside existing local DBs to bring the discoverability and connectivity, without replacing or altering the local solutions

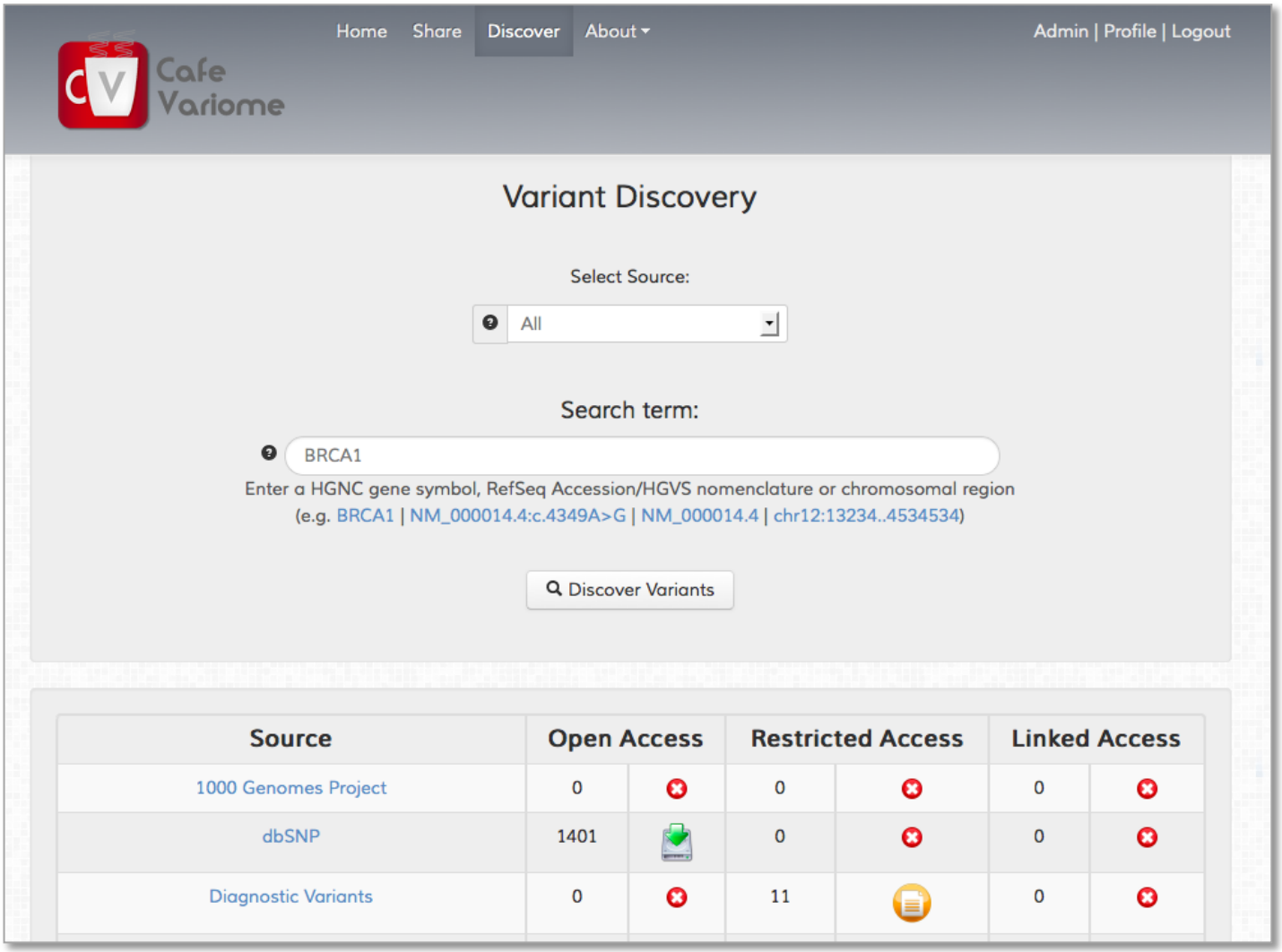
IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Café Variome (2)



IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Café Variome (3)



Café Variome (4)

Open Discovery – Reporting Existence of Variants in Sources

Open Access

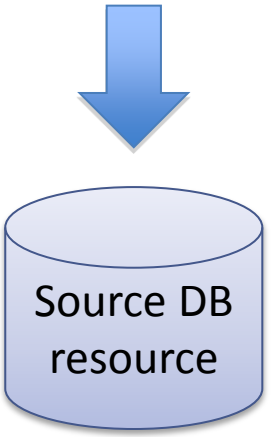
Core info for each variant record is shown & made available for download

Restricted Access

Core or full record details are provided per record, if:

- User is pre-approved by group access permissions set by data owner
- Access is approved after facilitated email request to the data owner

Linked Access



No data, only link to the data source is reported.

Access then control managed by source db

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

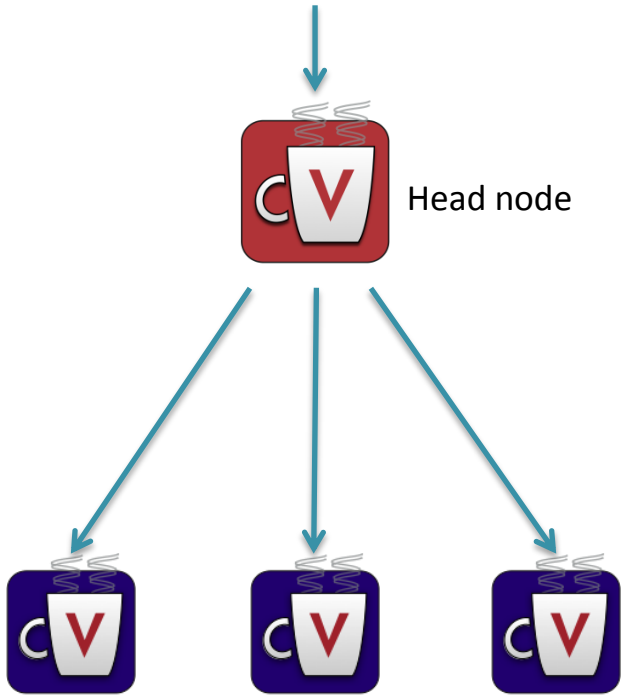
Café Variome (5)

Variant Report for vx62

Cafe Variome ID	vx62
Source link	-
Gene symbol / LRG	BRCA1 (LRG_292)
Reference sequence	NM_007294.3
HGVS description	c.4956G>A
Phenotype	Breast-ovarian cancer, familial, 1
Individual ID	-
Gender	-
Source/submitter ID	diagnostic
Genomic location	chr17:41222975-41222975 View in UCSC Genome Browser
Comment	Submitted from Gensearch tool during Cafe Variome pilot study

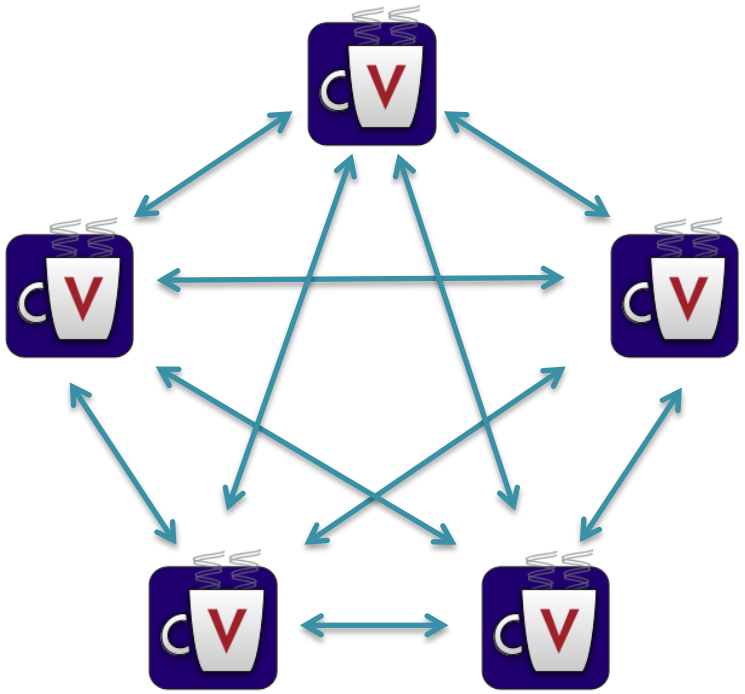
Café Variome (6)

External searches



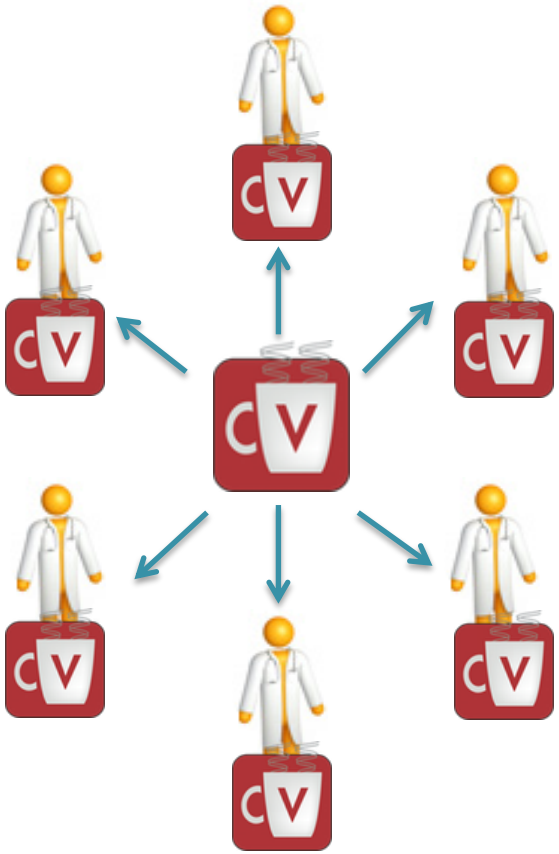
- Searches are performed through one nominated head node

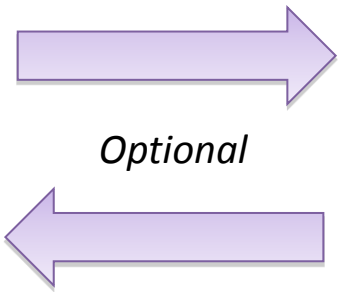
Internal searches

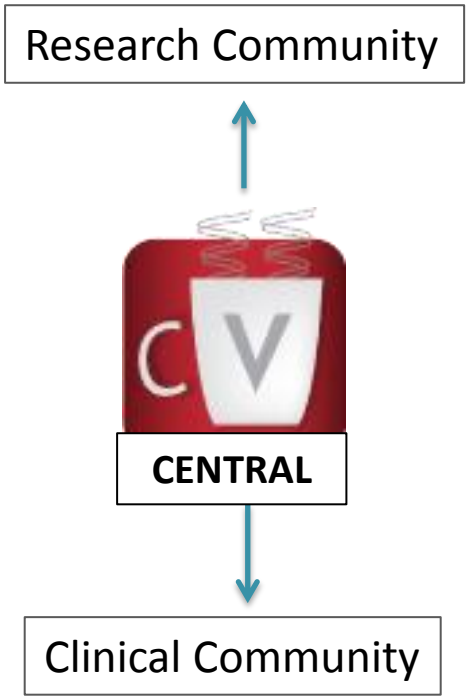


- Searches can be performed from any node in the network

Café Variome (7)







IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM

Café Variome (8)

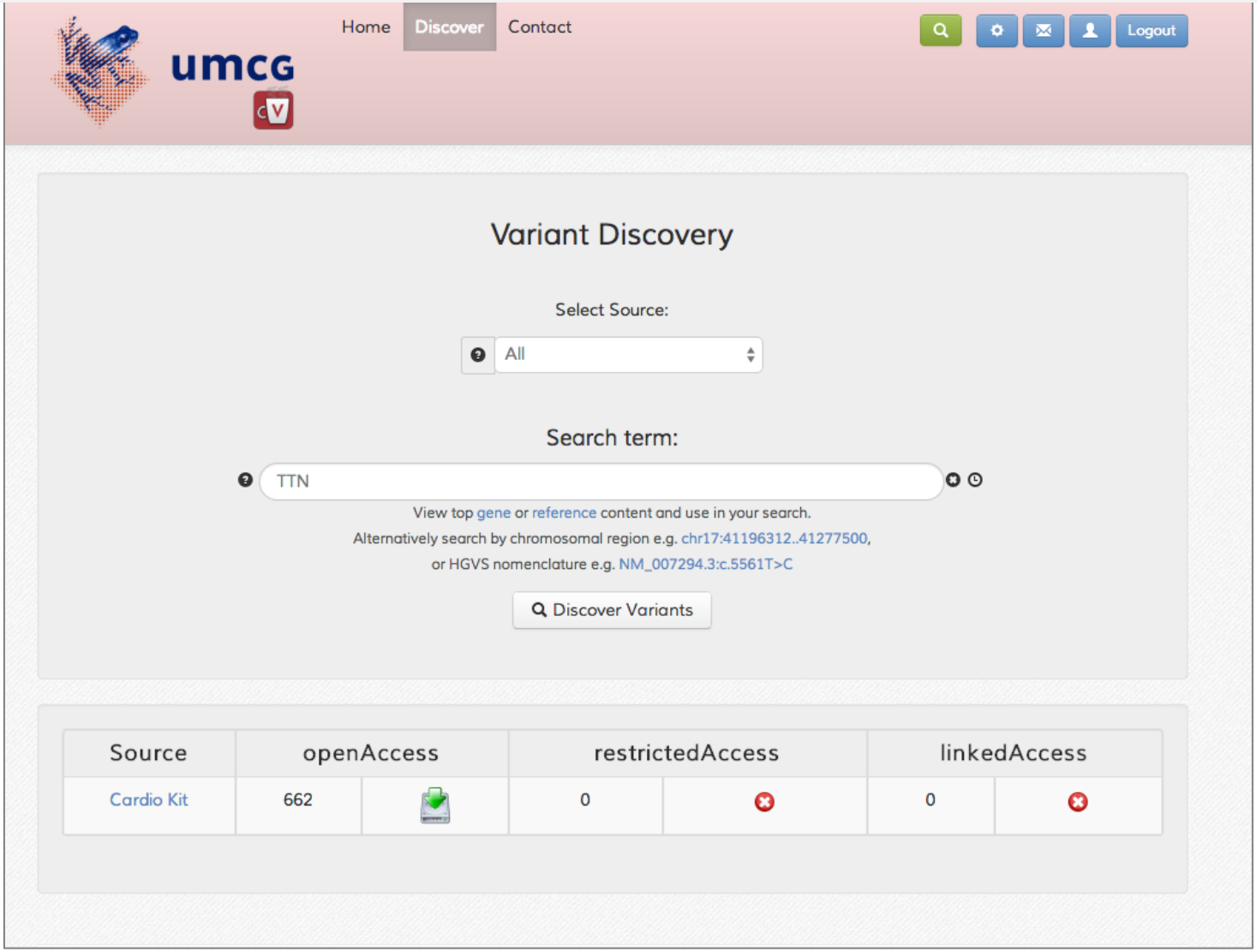
Disease consortia / diagnostic networks:

- Ehlers Danlos Syndrome Consortium (Ghent Uni, Belgium & others)
- Brugada Syndrome Network (lead by University of Girona, Spain)
- Sheffield Children's NHS Trust (Osteogenesis imperfecta)
- Congenital Adrenal Hyperplasia (lead by Medizinische Universität Innsbruck, Austria)
- Danish diagnostic network (lead by Aarhus University)
- Netherlands diagnostic network (lead by University of Groningen)

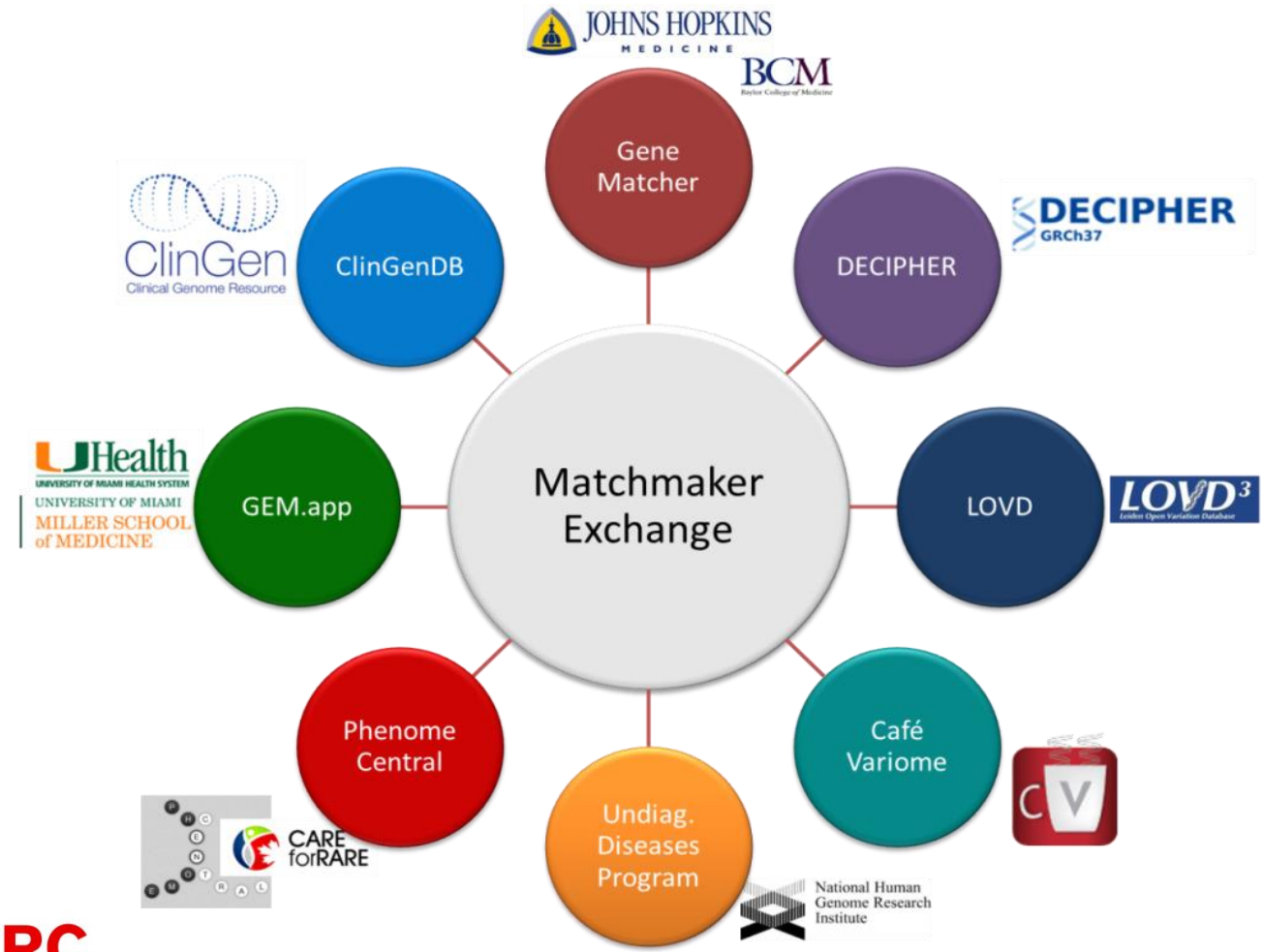
“Cafe Variome enabled” diagnostic software:

- Gensearch (Phenosystems)
- Alamut Visual (Interactive Biosoftware)

Café Variome (9)



Café Variome (10)



IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM