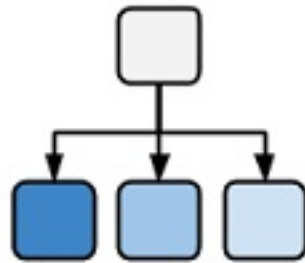


ONTOLOGIES IN COMPUTATIONAL BIOLOGY



Michel Dumontier, Ph.D.

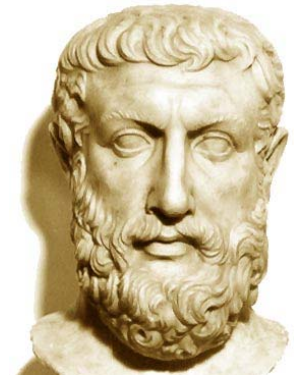
Distinguished Professor of Data Science
Maastricht University

Outline

- Introduction: what is an ontology and where can I find them
- Annotation: using ontologies for data annotation
- Analysis: gaining new insights in data through ontologies
- Reasoning: ontologies as formal knowledge bases

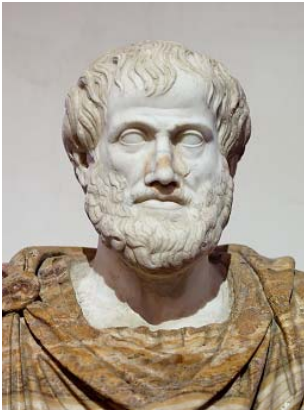
What is an ontology?

- Ontology stands for a ***logical discourse of existence***. It *aims to uncover and describe the nature and structure of existence*.
- Predominantly the domain of philosophy known as ***metaphysics***
- Address **questions** such as
 - What does it mean *to be*?
 - What constitutes the *identity* of an object?
 - What *categories* of things exist?
- Ontologies, when communicated to others, **foster a *shared understanding of things*** by making explicit a way to carve up the world



Greek philosopher Parmenides (515BC) proposed an ontological characterization of the fundamental nature of reality – akin to a grand unification theory

Early Bio-ontologists



Aristotle (384-322 BC)

- First systematic taxonomy of biology
- Classification of organisms by shared properties
- Used binomial *genus-differentia* nomenclature



Galen (130-210 AD)

- Systematic description of diseases, signs and symptoms.
- In *De Februm Differentia* description of fever symptoms he uses the Aristotelian *genus-differentia* approach

genus–differentia definitions are one way to specify ontologies

A type of *intensional* definition - where necessary and sufficient conditions are specified - composed of two parts:

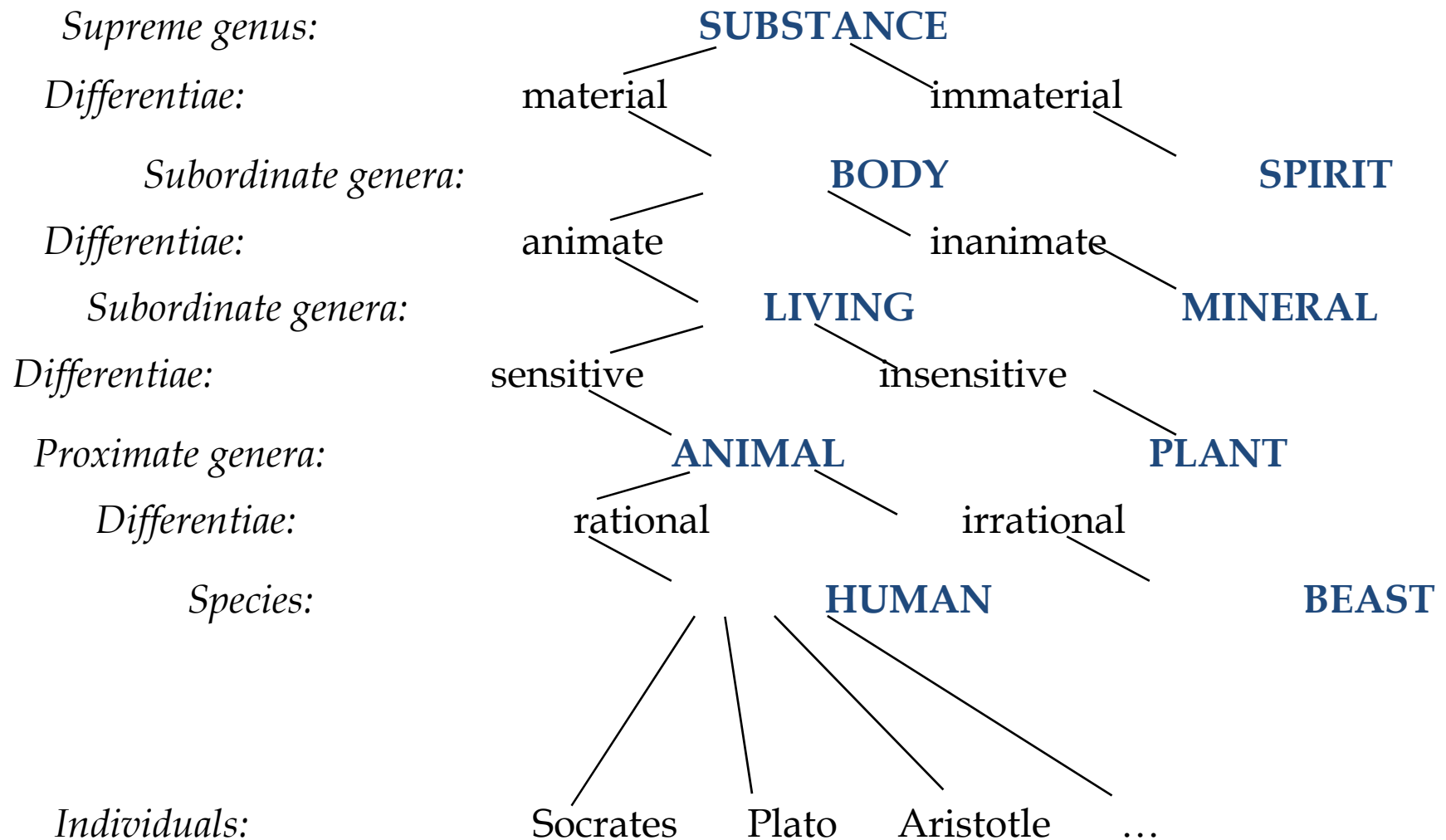
genus: Serves as the basis for a new definition; all definitions with the same genus are considered members of that genus.

differentia: The portion of the definition that is not provided by the genus.

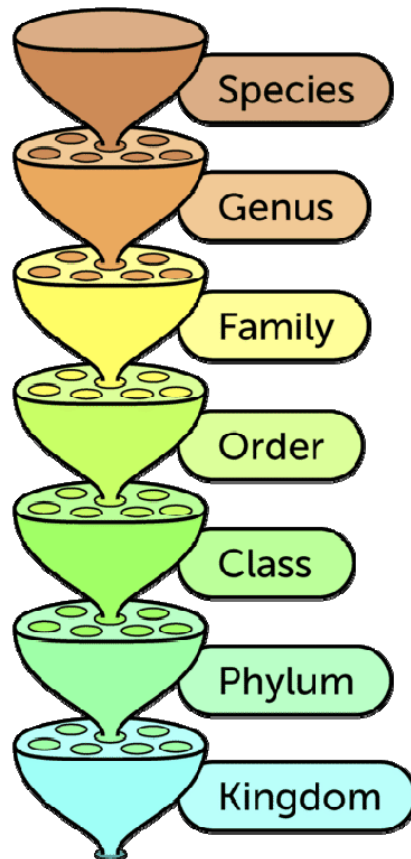
a **rhombus**: a **quadrilateral** that has bounding sides which all have the same length.

a **square**: a **rhombus** that has interior angles which are all right angles.

Porphyry's depiction of Aristotle's Categories



Biological Taxonomy



Homo sapiens

Members of the genus *Homo* with a high forehead and thin skull bones.

Homo

Hominids with upright posture and large brains.

Hominids

Primates with relatively flat faces and three-dimensional vision.

Primates

Mammals with collar bones and grasping fingers.

Mammals

Chordates with fur or hair and milk glands.

Chordates

Animals with a backbone.

Animals

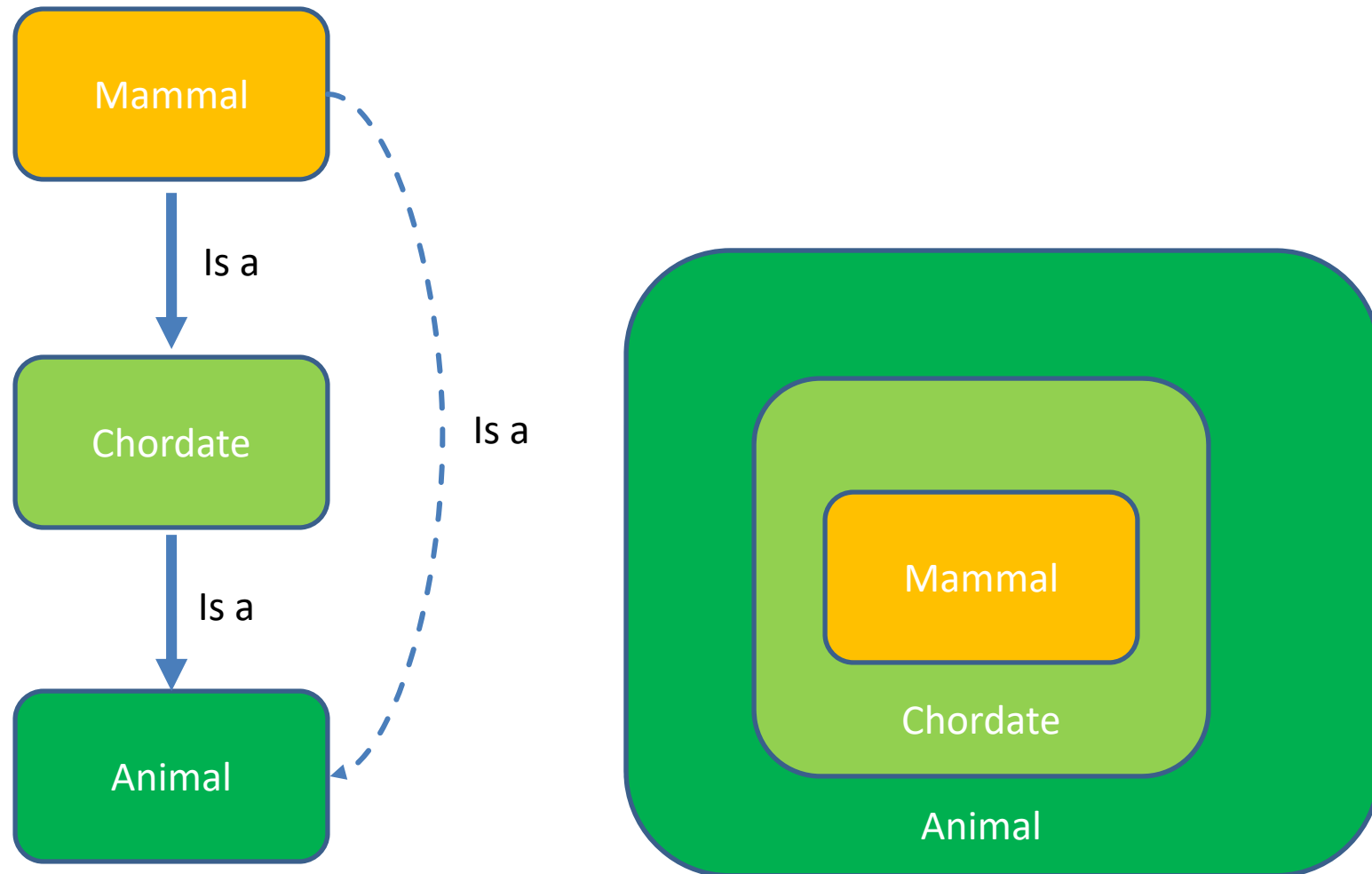
Organisms able to move on their own.

- A biological classification (taxonomy) by **Carl Linnaeus** in his *Systema Naturae* (1735)
- Three kingdoms, divided into classes, and they, in turn, into orders, families, genera, and species, with an additional rank lower than species.

*Rank: a classification
Of taxonomic categories*

*Biological taxonomy:
an is-a hierarchy
of biological types*

Genus-differentia illustrates basic inference vis-à-vis the “is a” relationship



Development of an *increasingly* applied notion of ontology

An explicit specification of a conceptualization

- Thomas Robert Gruber, 1993 (inventor of Siri)
- A **conceptualization** is the way we think about a domain (a “system of categories accounting for a particular view on the world”, i.e., a philosophical ontology)
- A **specification** provides a formal way of writing it down (and making it accessible to humans **and** machines)

*A **formal** specification of a **shared** conceptualization*

- Borst 1997

*An ontology specifies a **vocabulary** with which to make assertions, which may be inputs or outputs of knowledge agents (such as a software program). ... **an ontology must be formulated in some representation language***

- Gruber (2007)

An ontology is defined by *axioms* in a **formal language** with the goal to provide an unbiased (domain- and application-independent) view on reality

How is an ontology different than a...

- **Folksonomy**
 - A collection of terms (tags) to enhance categorization.
- **Glossary**
 - List of terms with definitions and explanations in natural language
- **Controlled Vocabulary**
 - An enumeration of terms defined to be shared and reused.
- **Hierarchy**
 - A nested set of terms
- **Taxonomy**
 - A hierarchy that uses the “is a” relation.
- **Meronomy**
 - A hierarchy that uses the “part of” relation.
- **Classification**
 - A set of categories in which objects are grouped into a hierarchy

Ontologies vs Classifications

Classifications focus on:

- Access based on predetermined criteria (encoded by syntactic keys)
- Uniform syntactic access to information

Ontologies focus on:

- Meaning of terms
- Nature and structure of a domain of discourse
- Shared understanding

Ontologies vs Knowledge Bases

Knowledge bases:

- Assertional component
 - Reflects specific states of affairs
 - Designed for problem solving
- Terminological component (ontology)
 - Independent of states of affairs
 - Designed to support terminological services
 - ... but are independent of the actual terminology used
- Ontological formulas are **invariant, necessary** information

Why develop an ontology?

- To provide a **formal specification** of biomedical knowledge
- To provide a **classification** of biomedical entities
- To develop a **common understanding** of the entities in a given domain
- To enable **reuse** of **data** and **knowledge**, e.g., when developing a database
- To enable biomedical **discovery**

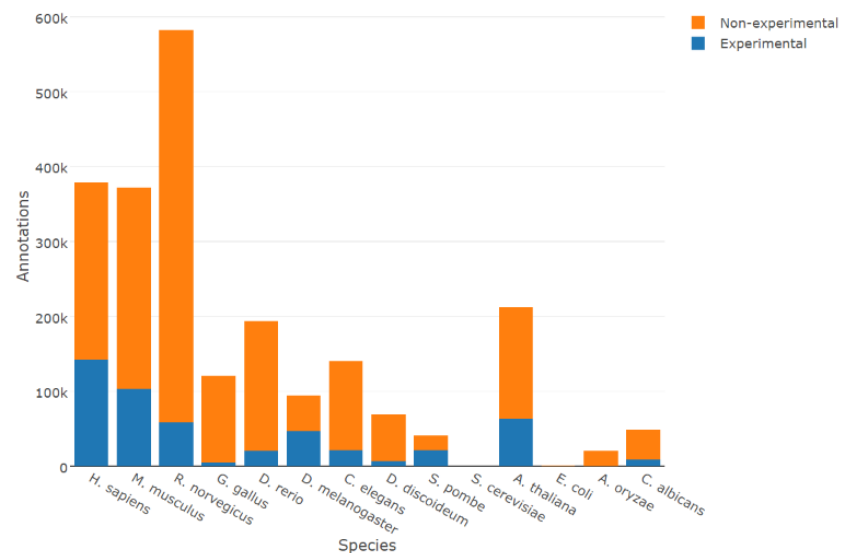
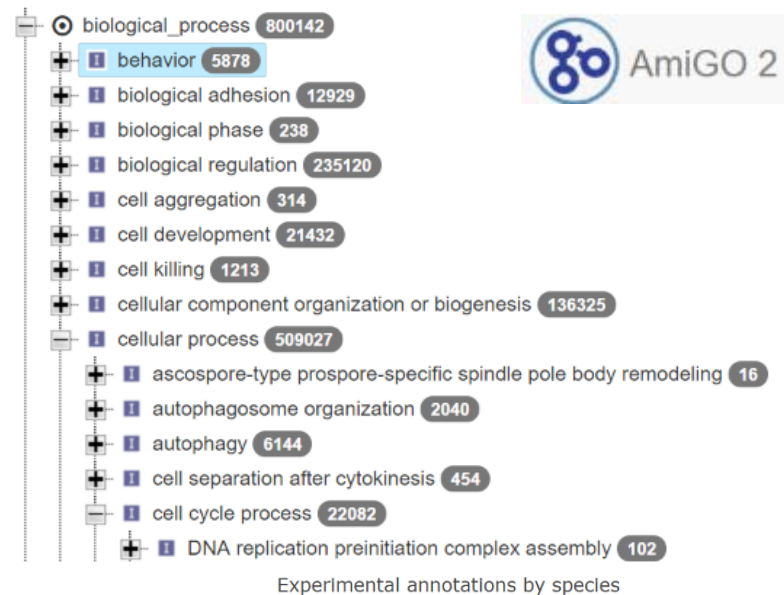
Gene Ontology

Arguably one of the most *successful* ontology projects in the life sciences.

Millions of annotations on hundreds of thousands of genes using GO terms.

The GO defines types used to describe gene function. It classifies functions along three aspects:

- **molecular function**
 - what gene products do
- **cellular component**
 - where gene products operate
- **biological process**
 - The pathways and processes that gene products participate in



UniProtKB - P34144 (RAC1A_DICDI)

Protein | Rho-related protein rac1A

Gene | rac1A




Organism | *Dictyostelium discoideum* (Slime mold)

Status |  Reviewed - Annotation score:  - Experimental evidence at protein levelⁱ

Functionⁱ

Overexpression promotes the formation of filopodia and membrane ruffles.  1 Publication ▼

Regions

Feature key	Position(s)	Length	Description
Nucleotide binding ⁱ	10 – 17	8	GTP  B
Nucleotide binding ⁱ	57 – 61	5	GTP  B
Nucleotide binding ⁱ	115 – 118	4	GTP  B

Manual assertion based on experiment inⁱ

"Rac1 GTPases control filopodia formation, cell motility, endocytosis, cytokinesis and development in Dictyostelium."

Dumontier M., Hoecht P., Mintert U., Faix J.
J. Cell Sci. 113:2253-2265(2000) [PubMed] [Europe PMC] [Abstract]

Cited for: INTERACTION WITH RGAA, FUNCTION.

GO - Molecular functionⁱ

- GTP binding  Source: UniProtKB-KW
- protein kinase binding  Source: dictyBase ▼

GO - Biological processⁱ

- positive regulation of actin filament polymerization  Source: dictyBase ▼
- Rac protein signal transduction  Source: dictyBase ▼

[Complete GO annotation...](#)

Keywords - Ligandⁱ

GTP-binding, Nucleotide-binding

GO facilitates interoperability of function descriptions across species

Term Information ?	
Accession	GO:0005525
Name	GTP binding
Ontology	molecular_function
Synonyms	None
Alternate IDs	None
Definition	Interacting selectively and non-covalently with GTP, guanosine triphosphate. <i>Source:</i> GOC:ai
Comment	None
History	See term history for GO:0005525 at QuickGO
Subset	gosubset_prok
Related	Link to all genes and gene products annotated to GTP binding.
	Link to all direct and indirect annotations to GTP binding.
	Link to all direct and indirect annotations download (limited to first 10,000) for GTP binding.

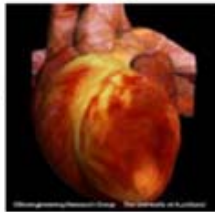
Organism	
+	- (8780) Eukaryota
+	- (5564) Metazoa
+	- (4905) Vertebrata
+	- (3269) Mammalia
+	- (1661) Fungi
+	- (1268) Danio rerio
+	- (1013) Rattus norvegicus
+	- (848) Viridiplantae
+	- (590) Bacteria
+	- (569) Mus musculus
+	- (455) Dictyostelium discoideum
+	- (447) Homo sapiens
+	- (433) Arabidopsis thaliana
+	- (387) Canis lupus familiaris
+	- (376) Bos taurus
+	- (346) Sus scrofa
+	- (332) Caenorhabditis elegans
+	- (307) Gallus gallus
+	- (209) Saccharomyces cerevisiae S288c
+	- (165) Drosophila melanogaster
+	- (115) Schizosaccharomyces pombe

Ontologies across scales

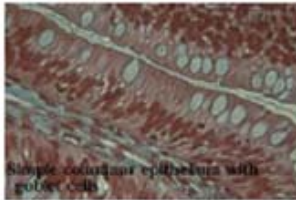
Organism



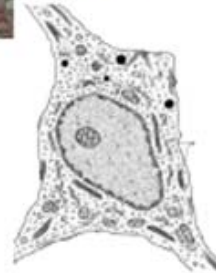
Organ



Tissue



Cell



Organelle



Protein



some disease and phenotype ontologies

- **Disease Ontology (DO)**
 - standardized ontology for human disease
 - Mapped to major terminologies, UMLS, MeSH, ICD10 etc.
 - 11,280 classes
- **Human Phenotype Ontology (HPO)**
 - phenotypic features encountered in human hereditary and other disease
 - 15,381 classes
- **Mammalian Phenotype Ontology (MP)**
 - Phenotypic features encountered in animal models
 - 12,805 classes
- **Experimental Factor ontology (EFO)**
 - application ontology
 - Imports classes from other phenotype and related ontologies (MIREOT)
 - 19,094 classes
- **Unified Medical Language System (UMLS)**
 - US National Library of Medicine
 - terminology, classification and coding standards
 - 8M normalized concepts
- **SNOMED-CT**
 - clinical terminology, diseases, diagnostics and procedures
 - 324,129 classes
- **NCI thesaurus**
 - vocabulary for clinical care, translational and basic research, and public information and administrative activities.
 - 118,941 classes
- **LOINC**
 - labs, vitals signs, clinical documents
 - 187,123 classes
- **ICD-10**
 - disease, epidemiology, billing
 - 12,450 classes

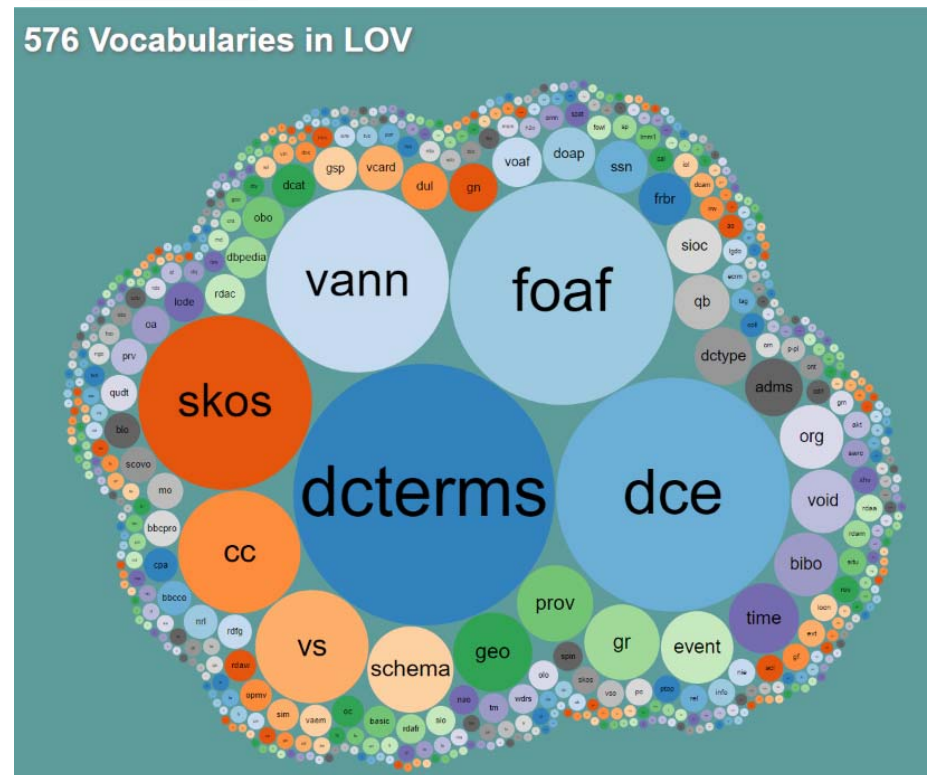
Where can we get ontologies?



Statistics	
Ontologies	517
Classes	7,801,718
Resources Indexed	48
Indexed Records	39,359,542
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932



Linked Open Vocabularies (LOV)



Formalization

- Formalization is the process by which we **map a conceptualization into a logical representation**.
- We *logically* combine the terms to form **expressions**, which have an *unambiguous* interpretation, and hence can be **automatically reasoned** about.

Logic-Based Ontologies

Can Be Constructed From *Concept and relation* Lego

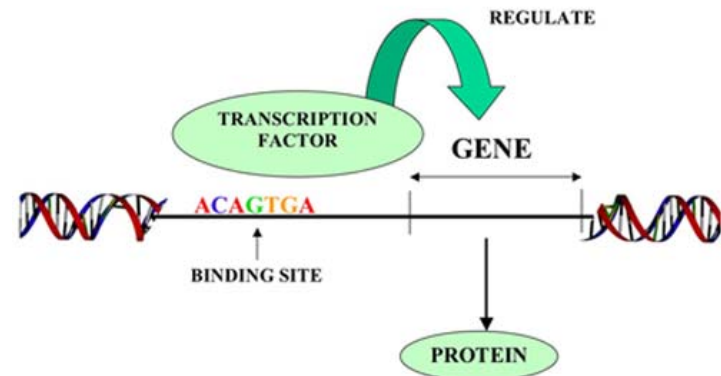


Have you heard of OWL?



Description logics offer the building blocks for constructing *computable* ontologies

A **transcription factor** is a protein that binds to DNA and regulates the rate of transcription



molecule ontology

'transcription factor'
equivalentTo
'protein'
that **'binds to'** some **DNA**
and **'regulates'** some **'rate of transcription'**



function ontology

The Web Ontology Language (OWL) Has Explicit Semantics



**It can be used to capture knowledge in a
machine understandable way**

OWL specifies a vocabulary and grammar to express more precisely what you mean

Enhanced vocabulary (strong axioms) to express knowledge relating to classes, properties, individuals and data values

- **Subclass, Equivalence**
- **Disjointness (disjointWith (classes), differentFrom (instances))**
- **Quantification (some, only, 0->n)**
 - existential, universal, cardinality restriction
- **Negation (not)**
- **Disjunction (or)**
- **property characteristics**
 - transitive, functional, inverse functional, symmetric, antisymmetric, reflexive, irreflexive
- **complex class expressions** in domain and range restrictions
- **property chains**

What is a class?

A **class** represents a **group of individuals** that share one or more things in common.

e.g. The class of “transcription factor” represents all entities that satisfy the criteria for class membership.

What is an instance?

An **instance** is an *individual* that is a member of a **class**.

e.g. An *individual* transcription molecule in the cell is an instance of the class “transcription factor”

e.g. Michel is an instance of Person

An individual can be a member of more than one class. For instance a transcription factor is also a member of the class Protein.

An individual may have properties

These properties can refer to attributes of the individual or relations to other **individual**.

That individual transcription factor *is bound* to a single molecule of DNA

Robert *is a colleague of* Michel

subClass

A **subClass** is a class that minimally exhibits all the attributes of a parent class, and potentially more.

Transcription Factor subClassOf Protein
all transcription factors are also proteins -> each transcription factor shares all the attributes that proteins do.

However, the inverse, that all proteins are transcription factors is **not** implied.

Equivalent class

Two classes are **equivalent** if they contain exactly the same individuals.

For example, we often use the term Protein and Polypeptide interchangeably, meaning that every instance of the class Protein is also an instance of class Polypeptide, and vice versa.

Disjoint classes

Two classes are **disjoint** if they contain exactly no same individuals. In some cases, membership in one class precludes membership in another.

For example, no molecule can both be wholly a DNA molecule and also be a Protein molecule.
`owl:disjointClasses (DNA, Protein)`

Same and different individuals

OWL does not assume that two names imply two different individuals. In fact, one individual may have two names – e.g. **P38398** and **Q7KYU9** are two names for **BRCA1** - so it would be wrong to conclude that these refer to two different individuals.

However, OWL provides a mechanism to explicitly indicate that individuals are the same as (owl:sameAs) or different from (owl:differentFrom) one another.

Same and different individuals

This has important implications when counting the number of entities in a knowledge base.

Individual: **P38398**

Individual: **Q7KYU9**

Individual: **Q16512**

-> how many individuals are there? 1

P38398 owl:sameAs **Q7KYU9**

-> how many individuals are there? 1

P38398 owl:differentFrom **Q16512**

-> how many individuals are there? 2

More complex class expressions

We can define more complex descriptions of a class by combining classes and properties together. The basic constructors are 'and'/'that', 'or', 'not', 'some', 'only', 'min', 'max'.

'transcription factor'

equivalentTo

'protein'

that 'binds to' some DNA

and 'regulates' min 1 'rate of transcription'

Reasoning over OWL ontologies

- **Consistency:** determines whether the ontology contains **contradictions**.
(1) DNA disjointFrom Protein (2) myprotein instanceOf protein (3) myprotein instanceOf DNA
- **Satisfiability:** determines whether classes can have **instances**.
(1) DNA disjointFrom Protein (2) DNA subClassOf Protein
- **Subsumption:** are **all instances** of one class **also instances** of another class?
(1) Protein subClassOf Molecule
- **Classification:** *repetitive* application of **subsumption** to discover implicit subclass links between named classes
(1) Transcription Factor equivalentTo Protein ...
-> (2) Transcription Factor subClassOf Protein
- **Realization:** find the most **specific** class that an individual belongs to.



Protégé's plug-in architecture can be adapted to build both simple and complex ontology-based applications. Developers can integrate the output of Protégé with rule systems or other problem solvers to construct a wide range of intelligent systems. Most important, the Stanford team and the vast Protégé community are here to help.



ACTIVE COMMUNITY

Protégé is actively supported by a strong community of users and developers that field questions, write documentation, and contribute plug-ins.



W3C STANDARDS SUPPORT

Protégé fully supports the latest OWL 2 Web Ontology Language and RDF specifications from the World Wide Web Consortium.



EXTENSIBLE OPEN SOURCE ENVIRONMENT

Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development.

DOWNLOAD NOW

USE WEBPROTÉGÉ

The screenshot shows the Protégé web interface for an ontology named 'untitled-ontology-31'. The left sidebar displays a class hierarchy with 'Protein' as a subclass of 'Organic Molecule'. The main area shows the 'Protein' class with its description: 'A protein is an organic molecule that is primarily composed of a linear chain of amino acids.' The right sidebar shows the 'Instances' tab, which is currently empty.



TopBraid Composer™ Maestro Edition



TopBraid Composer Maestro Edition™

Ready to build powerful semantic web and Linked Data applications for your enterprise? Let's get to work.

Buy Now >>

TopBraid Composer™ Maestro Edition (TBC-ME) combines world's leading semantic web modeling capabilities with the most comprehensive data conversion options and a powerful Integrated Development Environment (IDE) for building semantic web and Linked Data applications.

The screenshot shows the TopBraid Composer Maestro Edition IDE. The main window displays a complex ontology diagram with classes like 'LuxuryHotel', 'City', 'UrbanArea', and 'Destination'. The left sidebar shows a list of ontologies, including 'travel.owl', 'family.owl', and 'pizza.owl'. The right sidebar shows a list of properties, including 'hasAccommodation', 'hasActivity', and 'hasPart'.

untitled-ontology-31 (<http://www.semanticweb.org/micheld/ontologies/2016/9/untitled-ontology-31>) : [<http://www.semanticweb.org/micheld/ontologies/2>

File Edit View Reasoner Tools Refactor Window Debugger Ontop Help

< > untitled-ontology-31 (<http://www.semanticweb.org/micheld/ontologies/2016/9/untitled-ontology-31>)

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: Protein

Class Annotations Class Usage

Annotations: Protein

Annotations +

`rdfs:label` [language: en]
Protein

`rdfs:comment`
A protein is an organic molecule that is primarily composed of a linear chain of amino acids.

Description: Protein

Equivalent To +

SubClass Of +

- 'has part' some 'Amino Acid'
- 'Organic Molecule'

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

Disjoint With +

Disjoint Union Of +

Annotation property hierarchy Datatypes

Individuals by type

Object property hierarchy Data property hierarchy

Object property hierarchy: 'has part'

owl:topObjectProperty

- 'has part'

owl:Thing

- 'Organic Molecule'
- Protein
- 'Amino Acid'

*differentia
genus*

Protein

Class hierarchy Data restriction creator
Object restriction creator Class expression editor

'has part' some

{
Self
'Amino Acid'
'Organic Molecule'
Protein

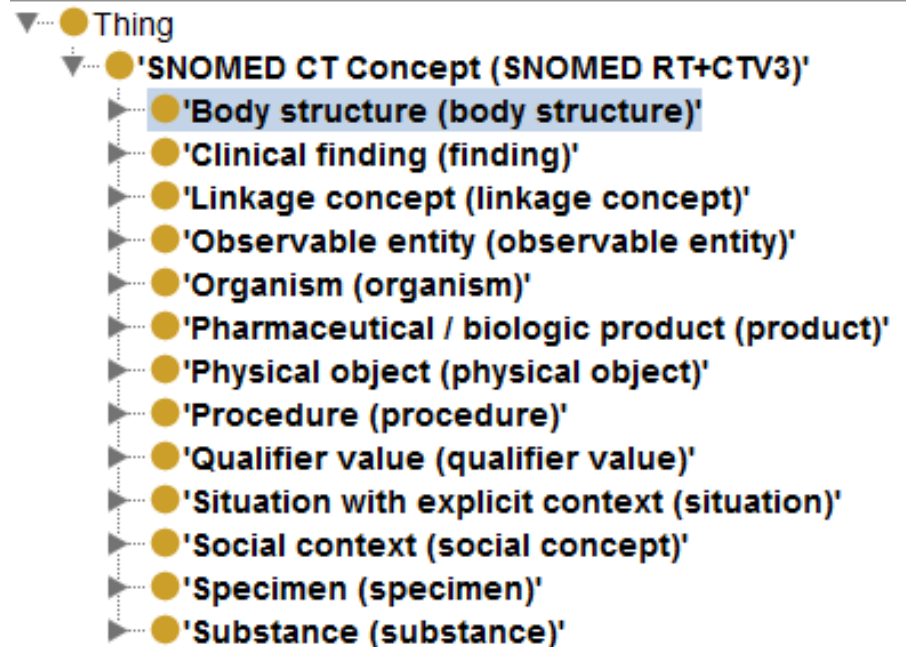
OK Cancel

Query answering

[illegible]


SNOMED-CT

- SNOMED-CT (Clinical Terms) ontology
- used in healthcare systems of more than 15 countries, including Australia, Canada, Denmark, Spain, Sweden and the UK
- used by major US providers
- ontology provides common vocabulary for recording clinical data
- 324,129 classes



SNOMED-CT

Description: 'Accessory breast (disorder)'

Equivalent classes 

● 'Congenital anomaly of breast (disorder)'
and 'Congenital malformation (morphologic abnormality)'
and (RoleGroup some
 (('Associated morphology (attribute)' some 'Supernumerary structure (morphologic abnormality)')
 and ('Finding site (attribute)' some 'Breast structure (body structure)'))))
and (RoleGroup some ('Occurrence (attribute)' some 'Congenital (qualifier value)'))

- Pattern based knowledge capture
- Requires some training and an information system to implement

SNOMED - Verification

- **Kaiser Permanente** extended SNOMED to express, e.g.:
 - *non-viral pneumonia* (**negation**)
 - *infectious pneumonia* is caused by a *virus* or a *bacterium* (**disjunction**)
 - *double pneumonia* occurs in *two lungs* (**cardinalities**)
- This is easy in **SNOMED-OWL**
 - but reasoner failed to find expected **subsumptions**, e.g., that *bacterial pneumonia* is a kind of *non-viral pneumonia*
- Ontology **under-constrained**: need to add **disjointness** axioms
 - *virus* and *bacterium* must be disjoint
- Adding **disjointness** led to **surprising results**
 - many classes become inconsistent, e.g., *percutaneous embolization of hepatic artery using fluoroscopy guidance*
- Cause of **inconsistencies** identified in the class *groin*
 - *groin* asserted to be subclass of both *abdomen* and *leg*
 - *abdomen* and *leg* are disjoint
 - modelling of *groin* (and other similar “junction” regions) identified as incorrect

QUICK PHENOTYPE SEARCH:

▼ BEHAVIOR, COGNITION AND DEVELOPMENT

☐ NA ☐ Y ☐ N Global developmental delay
☐ NA ☐ Y ☐ N Delayed fine motor development
☐ NA ☒ Y ☐ N **Delayed gross motor development**
☐ NA ☐ Y ☐ N Delayed speech and language development
☐ NA ☐ Y ☐ N Specific learning disability
Intellectual disability
☐ NA ☐ Y ☐ N Mild
☐ NA ☒ Y ☐ N **Moderate**
☐ NA ☐ Y ☐ N Severe
☐ NA ☐ Y ☒ N **Attention-deficit-hyperactivity disorder**
☐ NA ☐ Y ☐ N Autism
☐ NA ☐ Y ☐ N Behavioural/Psychiatric Abnormality

Other
(enter free text and choose among suggested ontology terms)

▼ NEUROLOGICAL

☐ NA ☐ Y ☐ N Generalized hypotonia
☐ NA ☐ Y ☐ N Seizures
☐ NA ☐ Y ☐ N Ataxia
☐ NA ☐ Y ☐ N Dystonia
☐ NA ☐ Y ☐ N Chorea
☐ NA ☒ Y ☐ N **Spasticity**
☐ NA ☐ Y ☒ N **Spinal-dysraphism**
☐ NA ☐ Y ☐ N Morphological abnormality of the central nervous system

Other
(enter free text and choose among suggested ontology terms)

▼ GROWTH PARAMETERS

Weight for age
☐ NA ☐ Y ☐ N <3rd
☐ NA ☐ Y ☐ N >97th
 Stature for age
☐ NA ☐ Y ☐ N <3rd
☐ NA ☐ Y ☐ N >97th
 Head circumference for age
☐ NA ☐ Y ☐ N <3rd
☐ NA ☐ Y ☐ N >97th
☐ NA ☐ Y ☐ N Hemihypertrophy

Other
(enter free text and choose among suggested ontology terms)

▼ CARDIAC

☐ NA ☒ Y ☐ N **Defect in the atrial septum**
☐ NA ☐ Y ☐ N Ventricular septal defect
☐ NA ☐ Y ☐ N Complete atrioventricular canal defect
☐ NA ☐ Y ☐ N Coarctation of aorta
☐ NA ☐ Y ☐ N Tetralogy of Fallot
☐ NA ☐ Y ☐ N Cardiomyopathy
☐ NA ☐ Y ☐ N Arrhythmia

CURRENT SELECTION

BEHAVIOR, COGNITION AND DEVELOPMENT

Delayed gross motor development Delete · Add details
 Intellectual disability, moderate Delete · Add details
 NO Attention deficit hyperactivity disorder Delete · Add details

NEUROLOGICAL

Spasticity Delete · Add details
 NO Spinal dysraphism Delete · Add details

CARDIAC

Defect in the atrial septum Delete · Clear details

Age of onset:

☐ Unknown
☒ Congenital onset
☐ Embryonal onset
☐ Fetal onset
☐ Neonatal onset
☐ Infantile onset
☐ Childhood onset
☐ Juvenile onset
☐ Adult onset
☐ Young adult onset
☐ Middle age onset
☐ Late onset

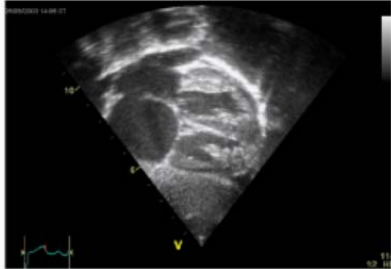
Pace of progression:

☒ Unknown
☐ Nonprogressive disorder
☐ Slow progression
☐ Progressive disorder
☐ Rapidly progressive
☐ Variable progression rate

Comments:

No complications

Image / photo (optional): + UPLOAD AND MANAGE



Medical report (optional): + UPLOAD AND MANAGE

None available

CRANIOFACIAL

NO Abnormal facial shape Delete · Add details

RESPIRATORY

NO Subglottic stenosis Delete · Add details

Phenotips

- Using controlled vocabulary (human phenotype ontology) for phenotyping
- Can collect demographics, medical history, family history, labs, findings

Girdea et al. (2013). Hum. Mutat., 34: 1057–1065. doi: 10.1002/humu.22347

PhenomeCentral: A Portal for Phenotypic and Genotypic Matchmaking of Patients with Rare Genetic Diseases

enter patient data → see similar patients → start a collaboration

A Quick phenotype search:
Enter keywords and choose from the suggested ontology terms

B Clinical symptoms and physical findings

GROWTH PARAMETERS
Head circumference for age
Microcephaly (<-3SD)

CRANIOFACIAL
Wide nasal bridge

EYE DEFECTS
Hypertelorism
NO Abnormal eye morphology

EAR DEFECTS
Hearing impairment

CARDIOVASCULAR
Ventricular septal defect

NEUROLOGICAL
Focal seizures

C LIST OF CANDIDATE GENES

#	GENE	OMIM: 600275	ENTREZ: 4853
1	NOTCH2		

GENECARDS: NOTCH2
REFSEQ: NM_024408
ENSEMBL: ENSG00000134250

D

Case ID	Diagnosis	Contact	Relevance
Undisclosed identifier	Undisclosed diagnosis	Undisclosed owner. Initiate anonymous contact	29%
Undisclosed identifier	Undisclosed diagnosis	Undisclosed owner. Initiate anonymous contact	24%
Undisclosed identifier	Undisclosed diagnosis	Undisclosed owner. Initiate anonymous contact	15%
Undisclosed identifier	Undisclosed diagnosis	Undisclosed owner. Initiate anonymous contact	14%
Undisclosed identifier	Undisclosed diagnosis	Undisclosed owner. Initiate anonymous contact	14%

E PHENOTYPIC FEATURES BREAKDOWN

ABNORMALITY OF THE VENTRICULAR SEPTUM	82%
The current patient (P0001152) presented with: Ventricular septal defect	The matched patient presented with: 1 undisclosed feature

ABNORMALITY OF SKULL SIZE	43%
The current patient (P0001152) presented with: Microcephaly	The matched patient presented with: 2 undisclosed features

ABNORMALITY OF THE NERVOUS SYSTEM	14%
The current patient (P0001152) presented with: Focal seizures	The matched patient presented with: 2 undisclosed features

F GENE MATCHING BREAKDOWN

VARIANT	ESTIMATED HARMFULNESS	VARIANT	ESTIMATED HARMFULNESS
chr1: 120611964 - 120611964 G → C (MISSENSE)	100%	Undisclosed position	97%
chr1: 120672572 - 120672572 C → T (MISSENSE)	97%	Undisclosed position	88%

G Contact a non-public case owner

1 Configure your message

SUBJECT
Interested in one of your non-public cases

Information about you:
☐ DISCLOSE YOUR NAME
☒ DISCLOSE YOUR EMAIL
☒ DISCLOSE YOUR MEMBERSHIP TO PHENOMECENTRAL GROUPS

Information about your case (P0001296):
☐ INCLUDE DIAGNOSIS INFORMATION
☐ INCLUDE A PHENOTYPE SUMMARY

Your requests:
☒ REQUEST MUTUAL VIEW ACCESS TO THE TWO SIMILAR CASES
If the recipient accepts, they gain view access to your case and you gain view access to theirs.
☒ REQUEST CONTACT INFORMATION

OTHER INFORMATION TO INCLUDE IN YOUR MESSAGE

2 Preview your message

This is the message the other user will receive:

SUBJECT
[PhenomeCentral] interested in one of your non-public cases

MESSAGE
Hello "undisclosed recipient name",
A PhenomeCentral user is interested in one of your non-public cases: "undisclosed case identifier". Please see their message below.
PhenomeCentral has identified significant similarities between one of your cases and one of mine.
I would like to grant you the rights to view my case and to obtain view access to your case, and to learn your contact information in order to further discuss these abnormalities with you.
To accept view privileges from this user and to grant them view access to "undisclosed case identifier", follow this link: "undisclosed URL".
Best wishes,
The PhenomeCentral team

SEND CANCEL

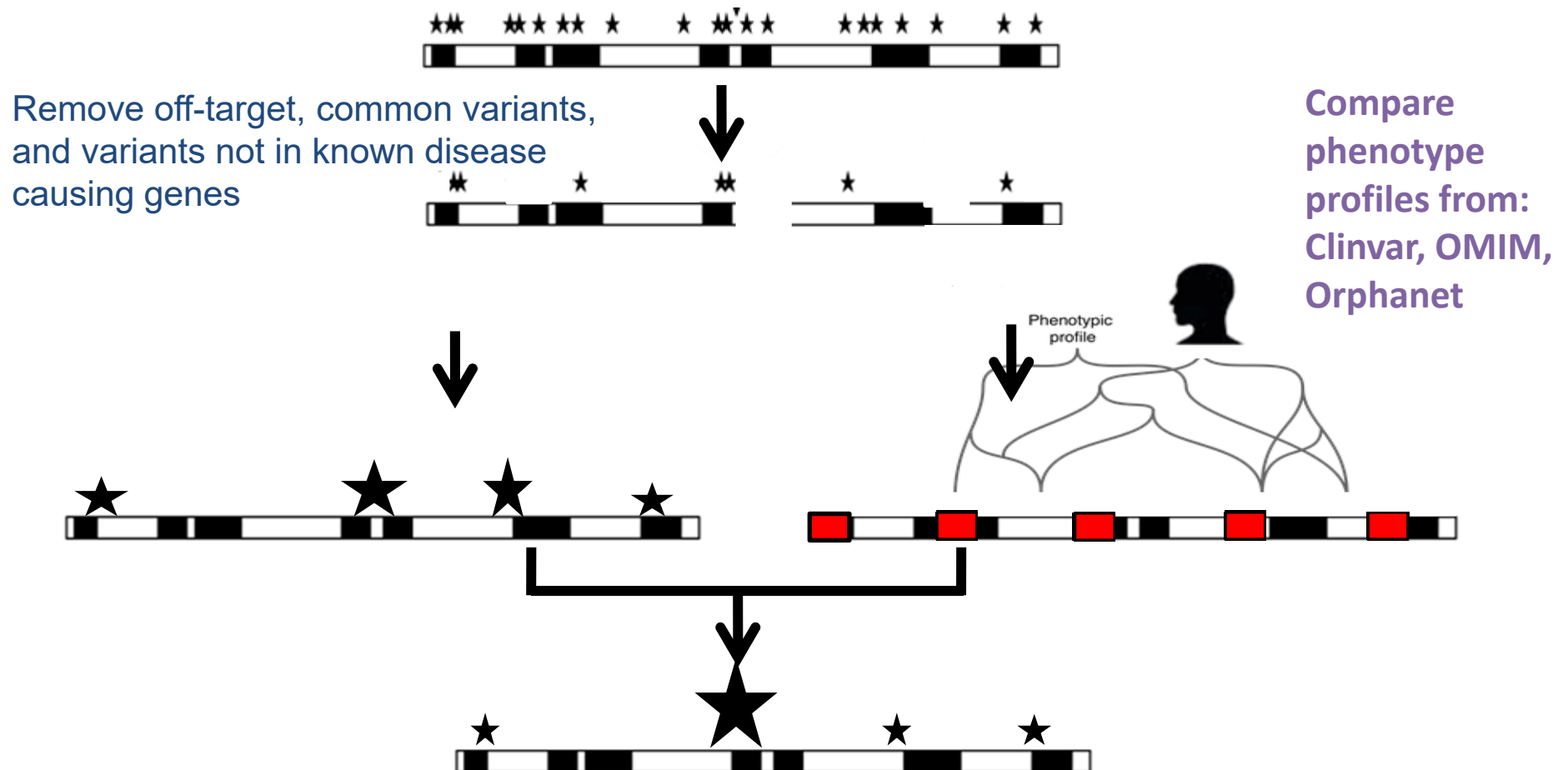
Human Mutation

Volume 36, Issue 10, pages 931-940, 31 AUG 2015 DOI: 10.1002/humu.22851

<http://onlinelibrary.wiley.com/doi/10.1002/humu.22851/full#humu22851-fig-0001>



Ontology-Aided Rare Disease Diagnosis



<http://compbio.charite.de/PhenIX/>

Credit: Damian Smedley & Will Bone
Zemojtel et al. Sci Transl Med. 2014. 6(252):252ra123

Summary

- Ontologies have a **rich history** in philosophy that has evolved to **modular and computable representation of human knowledge**
- **Description logics** (e.g. OWL) are the current favored formalism to build and test ontologies.
- Ontologies have a **variety of uses** from the answering questions to enabling sophisticated knowledge discovery.

Acknowledgements

Slides shamelessly adapted from *leaders* in the field:

- Mark Musen
- Barry Smith
- Ian Horrocks
- Robert Hoehndorf
- Melissa Haendel
- Paul Schofield